Agreement No. 22RD010

Impact of Air Pollution Exposure on Metabolic Health Outcomes for California Residents

Guangquan (Jason) Su, PhD, Principal Investigator
School of Public Health, University of California, Berkeley

Sadeer Al-Kindi, MD, Subcontract PI
Houston Methodist & Weill Cornell Medicine, Houston Methodist Research Institute

Michael Jerrett, PhD, Subcontract PI
Fielding School of Medicine, University of California, Los Angeles

Timothy T. Brown, PhD, Co-Investigator
School of Public Health, University of California, Berkeley

Rob Scot McConnell, MD, Project Advisor
Keck School of Medicine, University of Southern California

John Balmes, MD, Project Advisor
School of Medicine, University of California, San Francisco

October 2025
Prepared for the California Air Resources Board

Disclaimer

The statements and conclusions in this Report are those of the contractor and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

Acknowledgement

This report was prepared in fulfillment of CARB Agreement 22RD010, "Impact of Air Pollution Exposure on Metabolic Health Outcomes for California Residents," conducted by the University of California, Berkeley, in collaboration with Dr. Al-Kindi of the Houston Methodist DeBakey Heart & Vascular Center and Dr. Jerrett of the University of California, Los Angeles (UCLA). Work was completed as of October 2025.

We gratefully acknowledge the funding and support provided by CARB. We also acknowledge the invaluable data sources that made this study possible, including Type 2 diabetes incidence and medication use data from the California Health Interview Survey (CHIS) provided by the UCLA Center for Health Policy Research; emergency department (ED) visits and hospitalization data from the California Department of Health Care Access and Information (HCAI); and mortality data from the California Department of Public Health (CDPH).

We extend our sincere thanks to our project advisors, including Dr. Rob McConnell, MD (Keck School of Medicine, University of Southern California) and Dr. John Balmes, MD (School of Medicine, University of California, San Francisco), for their expert guidance and mentorship throughout the course of this study. Their insights and continued engagement were instrumental to the project's success.

List of Abbreviations

AOD – Aerosol Optical Depth
AQS – Air Quality System
ASPE – Office of the Assistant Secretary for Planning and Evaluation
BAT – Brown Adipose Tissue
BMI – Body Mass Index
CARB – California Air Resources Board
Caltrans – California Department of Transportation
CDC – Centers for Disease Control and Prevention
CDPH – California Department of Public Health
CHIS – California Health Interview Survey
CHPR – Center for Health Policy Research
CLR – conditional logistic regression
CTRL – Control areas
CI – Confidence Interval
CKD – Chronic Kidney Disease
CMAQ – Community Multiscale Air Quality Model
CNS – Central Nervous System
CO – Carbon Monoxide
CPHS – Committee for Protection of Human Subjects
CT – Census Tract
CV – Cardiovascular
DLNM – Distributed Lag Nonlinear Model
DM – Diabetes Mellitus
ED – Emergency Department
EOS – Earth Observing System
EPA – U.S. Environmental Protection Agency
ER – Endoplasmic Reticulum
ERF – Exposure Response Function
ESRI – Environmental Systems Research Institute
Exposome – Cumulative environmental, behavioral, and social exposures over the life course
FPL – Federal Poverty Level
GAM – Generalized Additive Model
GEOS-Chem – Goddard Earth Observing System–Chemistry Model
GLMs – Generalized Linear Models
GLUT4 – Glucose Transporter Type 4
GMCs – Goods Movement Corridors
GLM – Generalized Linear Model
HAP – Hazardous Air Pollutant
HCAI – California Department of Health Care Access and Information
HHS – U.S. Department of Health and Human Services

HIPAA – Health Insurance Portability and Accountability Act
HMO – Health Maintenance Organization
HMRI – Houston Methodist Research Institute
HR – Hazard Ratio
HPA – Hypothalamic–Pituitary–Adrenal (Axis)
ICD-9 – International Classification of Diseases, Ninth Revision
ICD-10 – International Classification of Diseases, Tenth Revision
IDF – International Diabetes Federation
IP - Inpatient
IPUMS NHGIS – Integrated Public Use Microdata Series, National Historical Geographic Information System
IQR – Interquartile Range
MR – Mean Ration in GLM modeling of LOSIRB – Institutional Review Board
LOS – Length of Stay
LUR – Land Use Regression
MAIAC – Multi-angle Implementation of Atmospheric Correction
MDA8 – Maximum Daily 8-hour Average
MODIS – Moderate Resolution Imaging Spectroradiometer
NASA – National Aeronautics and Space Administration
NAAQS – National Ambient Air Quality Standards
NGMCs – Non-Goods Movement Corridors
NHGIS – National Historical Geographic Information System
NLCD – National Land Cover Database
NO_2 – Nitrogen Dioxide
NTL – Nighttime Lights
NVDI – Normalized Difference Vegetation Index
O_3 – Ozone
OMB – Office of Management and Budget
OMI – Ozone Monitoring Instrument
OR – Odds Ratio
PAF – Population Attributable Fraction
PAH – Polycyclic Aromatic Hydrocarbon
Pb – Lead
PEP – Population Estimates Program
PIF – Potential Impact Fraction
PM_{10} – Coarse Particulate Matter (≤10 micrometers)
$PM_{2.5}$ – Fine Particulate Matter (≤2.5 micrometers)
ppb – parts per billion
R – Statistical Programming Language
RIA – Regulatory Impact Analysis
RNS – Reactive Nitrogen Species

ROS – Reactive Oxygen Species
RR – Relative Risk
SD – Standard Deviation
SE – Standard Error
SO_2 – Sulfur Dioxide
SRDC – Secure Research Data Center
T2D – Type 2 Diabetes Mellitus
Th1/M1 – T-helper Type 1 / Macrophage Type 1 (immune phenotype)
UCB – University of California, Berkeley
UCLA – University of California, Los Angeles
USC – University of Southern California
USD – United States Dollar
USGS – United States Geological Survey
VOC – Volatile Organic Compound
VSL – Value of a Statistical Life
VKT – Vehicle Kilometers Traveled
ZIP – Zone Improvement Plan (typically three or five digits)

Table of Contents:

List of Tables

List of Figures

Project Summary/Abstract

This study evaluated the effects of ambient air pollution on type 2 diabetes (T2D)–related outcomes across the disease continuum, from onset to mortality, using three statewide data sources in California: the California Health Interview Survey (CHIS), the Department of Health Care Access and Information (HCAI) hospital and emergency department (ED) discharge data, and mortality data from the California Department of Public Health (CDPH). The analysis covered 2010–2019 and focused on three major pollutants—nitrogen dioxide (NO_2), fine particulate matter ($PM_{2.5}$), and ozone (O_3).

Using CHIS data, we assessed both the incidence of T2D and the use of diabetes medications among adults. Higher annual exposures to air pollution were consistently associated with elevated odds of newly reported diabetes and greater medication use. For T2D incidence, NO_2, $PM_{2.5}$, and O_3 all showed positive associations across lag years, with the strongest and most consistent effects observed for $PM_{2.5}$ and O_3 ($p < 0.001$). For medication use, odds ratios (ORs) for NO_2 remained stable from lag 0 to lag 3 years (≈1.018) per interquartile change (IQR), while O_3 showed a slight increase (≈1.034 to 1.036) and $PM_{2.5}$ exhibited a slightly higher - rise (≈1.070 to 1.073), per IQR increase of respective exposure. These findings suggest that exposure to criteria air pollutants may both trigger diabetes onset and worsen disease control among those already diagnosed.

In addition to criteria pollutants, we evaluated five ambient air toxics—benzene, chromium, nickel, lead, and 1,3-butadiene—using annually aggregated exposure models that explained 59–90% of adjusted variance. Across California, long-term exposure to each toxic was consistently and positively associated with both incident T2D and diabetes medication use. In pooled analyses, incident T2D increased by 18% for chromium (OR≈1.18), 21% for lead (OR≈1.21), 31–33% for benzene (OR≈1.33), 23–30% for 1,3-butadiene (OR≈1.23–1.30), and 64–66% for nickel (OR≈1.64–1.66). Medication use showed parallel associations, ranging from 7% (chromium) to 32% (nickel) higher odds across short-term lags. These results identify specific air toxics as robust contributors to diabetes risk and management burden, beyond conventional traffic-related pollutants.

Analyses based on HCAI data revealed similar patterns for acute healthcare utilization. Short-term increases in NO_2 and $PM_{2.5}$ were associated with higher risks of T2D-related ED visits and hospital admissions. Among hospitalized patients, pollutant exposure was also linked to longer length of stay (LOS), a continuous outcome, indicating greater clinical severity. $PM_{2.5}$ showed the strongest associations, with risks elevated across all lag periods, while NO_2 demonstrated smaller but consistent effects. Transient effects were observed, with associations slightly attenuating from lag 0 to lag 3 days. O_3 exhibited mixed associations, with some analyses showing modest positive effects while others indicated weaker or nonsignificant relationships. Stratified analyses by demographic and socioeconomic characteristics (including

race-ethnicity, age, and insurance type) confirmed that these associations were robust across groups, with minority populations showing greater vulnerability.

Mortality analyses using CDPH data further confirmed the adverse impacts of air pollution on diabetes outcomes. Higher annual average NO_2 and $PM_{2.5}$ exposures in the 12 months prior to date of death were significantly associated with increased diabetes-related deaths, consistent with their roles in disease progression and systemic inflammation.

Taken together, findings across three separate datasets demonstrate that exposure to ambient air pollutants, particularly $PM_{2.5}$, NO_2 and air toxics, contributes to the onset, exacerbation, and progression of T2D. These results underscore the broad public health importance of air quality improvement in reducing the burden of diabetes in California. Corresponding health economic analyses indicate that reducing ambient concentrations of these pollutants could yield substantial economic benefits, including billions of dollars annually in avoided medical expenditures and billions of dollars in terms of the value of statistical lives saved. Reducing emissions from traffic and other pollution sources could yield substantial health benefits, particularly for populations already at elevated risk for diabetes and related complications.

Lay-Person Summary

Type 2 diabetes is a serious and growing health problem in California. While diet, physical activity, and genetics are well-known risk factors, this study shows that air pollution also plays an important role in the development and worsening of diabetes. Using large, statewide health data from California between 2010 and 2019, this study examined how exposure to common air pollutants, including traffic-related gases, fine particles in the air, ozone, and several toxic air pollutants, affects people across the full course of diabetes, from first diagnosis to medication use, emergency room visits, hospitalizations, and death.

The results consistently showed that people living in areas with higher air pollution were more likely to develop diabetes, more likely to require diabetes medications, and more likely to experience serious complications requiring emergency care or hospitalization. Higher pollution levels were also linked to longer hospital stays and increased risk of death from diabetes-related causes. Fine particulate matter, which comes largely from traffic, industry, and wildfire smoke, showed the strongest and most consistent health effects.

Importantly, the study found that not all populations are affected equally. Communities already facing social and economic disadvantages, including racial and ethnic minority groups and people without stable health insurance, experienced greater health impacts from the same pollution levels, highlighting existing environmental and health inequities across California.

Taken together, these findings show that air pollution is not just a respiratory or cardiovascular issue, it also worsens diabetes and increases the burden on individuals, families, and the healthcare system. Improving air quality, especially in heavily polluted and underserved communities, could help prevent new cases of diabetes, reduce complications for those already living with the disease, and lower healthcare costs statewide. Substantial economic benefits could be achieved, including billions of dollars annually in avoided medical expenditures and billions of dollars in terms of the value of lives saved. This study underscores the importance of clean air policies as part of long-term strategies to protect public health and reduce chronic disease in California.

Executive Summary

a. Background

Type 2 diabetes (T2D) continues to be a major and growing public health concern in California and worldwide. Beyond traditional risk factors such as obesity, age, and genetics, emerging research has increasingly identified environmental exposures — particularly air pollution — as important contributors to diabetes incidence, complications, and mortality. Despite accumulating evidence linking fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2) to metabolic and cardiovascular outcomes, limited work has comprehensively assessed their effects on multiple diabetes-related health outcomes using diverse population-level data sources across time. This project aimed to address these gaps by integrating statewide health datasets to evaluate the short- and long-term impacts of air pollution on T2D-related outcomes.

b. Objective

The study sought to examine the associations between ambient air pollutants (NO_2, $PM_{2.5}$, and O_3) and multiple T2D-related outcomes, including disease incidence, medication use, emergency department (ED) visits, hospital admissions, hospital length of stay (LOS), and mortality. Specific goals were to assess (1) whether short- and long-term exposures to air pollution increased risks for T2D health endpoints, (2) whether these effects varied across demographic and socioeconomic groups, and (3) how pollutant-specific lag patterns reflected the temporal nature of exposure-response relationships.

c. Methods

Analyses were conducted using three California-representative health datasets from 2010–2019, combined with high-resolution spatiotemporal exposure assessment, representing one of the first statewide evaluations of diabetes-related health outcomes using linked population-level health and advanced air pollution modeling data.

CHIS (California Health Interview Survey) provided individual-level data to model T2D incidence and medication use, both treated as a binary outcome. Logistic regression models were used to estimate associations with annual exposure to criteria pollutants (NO_2, $PM_{2.5}$ and O_3) and five air toxics (benzene, chromium, nickel, lead, 1,3-butadiene) adjusting for demographic, socioeconomic, and behavioral covariates.

HCAI (California Department of Health Care Access and Information) hospital discharge and ED datasets were used to assess short-term effects of daily pollutant variations on acute healthcare utilization, including ED visits, hospital admissions (binary outcome), and LOS (continuous outcome). Conditional logistic regression with distributed lag models (lags 0–3 days) was applied within a case-crossover framework, in which each case was self-matched to

control days occurring on the same day of the week one to four weeks prior to the event.. Stratified analyses by race-ethnicity, insurance type, region, sex, and age were also conducted.

CDPH (California Department of Public Health) death records were analyzed to estimate the associations between annual pollutant exposure and diabetes-related mortality using conditional logistic regression approaches, in which each diabetes-related death was matched to controls drawn from the same CDPH population who were still alive at the time of death, based on year and month of birth, sex, and race–ethnicity..

Pollutant exposure metrics were assigned using data derived from PI's modeled fields, and all analyses incorporated population weighting or covariate adjustment to minimize bias.

d. Results

Across the multiple statewide health datasets analyzed, we observed consistent evidence that ambient air pollution was associated with elevated risk of adverse T2D-related outcomes. While effect magnitudes and temporal patterns varied across pollutants and endpoints, the overarching signal was robust: short- and long-term exposures to several criteria pollutants and air toxics were linked to increased incidence, medication use, acute care utilization, and mortality among individuals with or at risk for diabetes.

CHIS data showed that long-term exposure to NO_2, $PM_{2.5}$, and O_3 was significantly associated with increased T2D medication use and incidence. While NO_2 effects remained stable across lag years, $PM_{2.5}$ exhibited the steepest increase in risk from lag 0 to lag 3, suggesting greater toxicity. O_3 effects were smaller but consistently positive across lags, indicating a modest contribution to T2D morbidity. Lags reflect single-day lag effects (e.g., lag0 for the same day/year, lag3 for three days/years before).

Air toxics analyses represent a novel component of this project. We found statistically significant and temporally consistent associations between long-term exposure to benzene, chromium, nickel, lead, and 1,3-butadiene and T2D CHIS outcomes statewide. In pooled incidence models, nickel exhibited the largest effect (approximately 64–66% higher odds), followed by benzene (31–33%), 1,3-butadiene (23–30%), lead (21%), and chromium (18%). These associations were stable across lag structures (0–3 years) and generally strengthened in mid-2000 diagnosis cohorts. Parallel findings for medication use (7–32% higher odds depending on pollutant) indicate that air toxics are associated not only with disease onset but also with greater treatment intensity. The magnitude, internal consistency, and replication across outcomes underscore the metabolic relevance of these hazardous air pollutants.

HCAI data revealed robust and coherent patterns for acute healthcare utilization. Increases in NO_2 and $PM_{2.5}$ were significantly associated with elevated risks of T2D-related ED visits and hospital admissions, as well as longer LOS among hospitalized patients, implying greater disease severity. $PM_{2.5}$ effects were generally the strongest, with NO_2 showing smaller but steady effects across lag days. A modest attenuation from lag 0 to lag 3 was observed, indicating transient but clinically relevant effects within a few days of exposure. In contrast, O_3 associations were mixed. While small positive associations were detected for hospital admissions and LOS at lag 0, inverse or null effects appeared at later lags, suggesting differential mechanisms or potential confounding by seasonal and photochemical factors. Stratified analyses confirmed that minority populations, especially Hispanic and Black individuals, tended to experience greater pollutant-related risks.

CDPH mortality data further supported the long-term health burden of air pollution, showing statistically significant increases in diabetes-related mortality with higher annual mean concentrations of $PM_{2.5}$ and NO_2. These findings align with national evidence linking fine particles and traffic-related pollutants to chronic metabolic and vascular stress.

Health economic analyses further demonstrated that reductions in ambient air pollution would yield substantial economic benefits related to Type 2 diabetes. Based on modeled changes in diabetes incidence, medication use, emergency department visits, hospitalizations, and mortality, lowering pollutant concentrations by **one interquartile range** of their observed distributions was associated with large avoidable costs across California. Annual medical expenditure savings were estimated to be approximately $1.4 billion for $PM_{2.5}$, $245 million for NO_2, and $1.2 billion for O_3. In addition, reductions in long-term exposure to $PM_{2.5}$ and NO_2 were associated with sizable avoided losses in the value of statistical life, totaling several billion dollars annually.

Overall, the consistency of findings across three separate datasets underscores a robust and causal relationship between ambient air pollution and multiple stages of T2D progression—from disease onset and treatment dependence to acute exacerbations and death.

e. Conclusion

This integrated statewide analysis demonstrates that air pollution, particularly $PM_{2.5}$, NO_2 and air toxics has a significant and measurable impact on T2D health outcomes across clinical severity levels. The effects were strongest for $PM_{2.5}$, modest but stable for NO_2, and mixed for O_3. Short-term exposures primarily influenced acute healthcare utilization, while long-term exposures contributed to disease development and mortality. The transient lag structure suggests that pollutant-induced exacerbations occur rapidly following exposure, emphasizing the importance of timely public health interventions during pollution peaks. Importantly, the

observed health impacts translate into substantial economic consequences. These economic benefits reinforce the public health relevance of air quality regulations, demonstrating that improvements in air pollution can reduce healthcare system burden.

These findings reinforce the need for continued efforts to reduce air pollution exposure, especially in vulnerable communities disproportionately affected by both diabetes and environmental burden. The combined epidemiologic and economic evidence demonstrates that air quality improvements reduce the health and healthcare burden of Type 2 diabetes in California. Policymakers, healthcare providers, and environmental agencies should consider integrating air quality control with chronic disease prevention strategies to mitigate the growing public health impact of diabetes in California.

Background

Type 2 diabetes (T2D) has emerged as one of the most prevalent and costly chronic diseases worldwide, imposing a growing burden on healthcare systems and communities. Characterized by insulin resistance and impaired glucose regulation, T2D contributes to cardiovascular disease, renal complications, neuropathy, and premature death.[1–4] In the United States, diabetes affects over 38 million people, with nearly 95% of cases classified as T2D.[5] California, the most populous and environmentally diverse state, bears a particularly heavy burden. Data from the California Health Interview Survey (CHIS) show that approximately 8% adults in the state has been diagnosed with diabetes (see Table 12), with prevalence disproportionately higher among Hispanic, Black, and lower-income populations. Although established risk factors such as age, obesity, diet, physical inactivity, and family history remain central to T2D development, accumulating evidence points to air pollution as an important, modifiable environmental determinant.[6–10] Epidemiologic and toxicologic studies over the past two decades have demonstrated that exposure to airborne pollutants, especially fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2), can adversely affect metabolic function.[10] These pollutants induce systemic inflammation, oxidative stress, and endothelial dysfunction, which are key pathways in insulin resistance and glucose dysregulation.[11,12] Long-term exposure to $PM_{2.5}$ has been linked to increased diabetes incidence and mortality,[10,13–15] while NO_2, a marker of traffic-related pollution, has also been associated with higher risk of diabetes onset and hospital admissions related to metabolic disorders.[16–18] The role of ozone (O_3) in diabetes-related outcomes remains less consistent. As a secondary pollutant formed through photochemical reactions between nitrogen oxides and volatile organic compounds, O_3 is well known for its respiratory toxicity.[19] However, its systemic metabolic impacts are complex. Some studies report that O_3 exposure may trigger oxidative stress and systemic inflammation, exacerbating insulin resistance,[20–22] while others find weaker or even inverse relationships.[23–26] These inconsistencies may arise from confounding by seasonality, temperature, or spatial averaging of exposure estimates.

Importantly, many prior studies examining air pollution and diabetes have relied on relatively coarse spatial resolution, often coarser than 1 km,[27–29] which may not adequately capture local variation in pollution exposure, especially in urban areas where concentrations vary sharply over short distances. Such exposure misclassification can bias risk estimates toward the null, obscuring true associations. To address this limitation, the present study leverages newly developed high-resolution air pollution exposure surfaces that provide daily concentrations of $PM_{2.5}$, NO_2, and O_3 at a 100 m grid resolution across California for more than thirty years (1989–2021).[30] These fine-scale surfaces integrate satellite observations, ground monitoring, meteorological data, and built environment characteristics to generate spatially and temporally resolved estimates that more accurately reflect individual and community-level exposures.

California offers a unique natural laboratory for this investigation, given its environmental diversity, persistent air quality challenges, and comprehensive health data infrastructure. Despite

major progress in air quality control, many regions, particularly in Central Valley and urban basins, continue to experience pollutant levels exceeding state and federal standards. Moreover, air pollution exposures and health burdens are unequally distributed: low-income and minority communities often experience higher exposure levels and limited access to healthcare, contributing to health disparities. Understanding how chronic and short-term air pollution exposures influence T2D outcomes in this context is essential for advancing environmental health equity and guiding evidence-based public health policies.

Previous research has typically focused on isolated health outcomes, such as diabetes incidence or mortality, and often relied on single datasets or limited geographic coverage. Few studies have integrated multiple statewide databases to investigate the full continuum of diabetes-related outcomes, from disease onset and medication use to acute care utilization and mortality. To fill this gap, this project combines three major California datasets to examine how both long-term and short-term air pollution exposures affect T2D outcomes:

- CHIS data is used to evaluate long-term annual associations with T2D incidence and medication use, reflecting disease development and management in the community.
- Healthcare Access and Information (HCAI) data is used to assess short-term daily associations with emergency department (ED) visits, hospital admissions, and length of stay (LOS), capturing acute exacerbations and healthcare utilization patterns.
- California Department of Public Health (CDPH) mortality data is used to investigate long-term annual associations with diabetes-related deaths, representing the ultimate burden of chronic exposure.

By integrating these complementary data sources with high-resolution exposure assessment, this project provides one of the most comprehensive evaluations to date of air pollution's impact on T2D in California. The study not only quantifies pollutant-specific risks across a spectrum of outcomes but also examines population heterogeneity by demographic and socioeconomic factors. The findings are expected to inform public health strategies, healthcare planning, and air quality regulations aimed at reducing environmental health disparities and mitigating the burden of diabetes statewide.

Study Objectives

The objective of this study is to comprehensively evaluate the impacts of ambient air pollution on T2D outcomes across California by examining multiple stages of disease development, progression, and severity. Specifically, the study aims to quantify associations between long-term and short-term exposure to key ambient air pollutants and diabetes-related outcomes, including disease incidence, medication use, acute healthcare utilization, and mortality. By leveraging high-resolution spatial and temporal exposure estimates, the study seeks

to reduce exposure misclassification and better characterize pollutant-specific effects across different exposure windows. In addition, the study aims to assess how these associations vary across demographic, socioeconomic, and geographic subgroups, thereby identifying populations that may be disproportionately affected by air pollution–related diabetes risks. Through the integration of multiple statewide health datasets, this work is intended to provide a comprehensive evidence base to support air quality management, public health planning, and policies aimed at reducing the burden of diabetes and advancing environmental health equity in California. Finally, the study seeks to estimate the potential healthcare utilization and economic impacts associated with air pollution–related diabetes outcomes, providing insight into the broader public health and economic benefits of improved air quality.

Task 1. Literature Review

T2D: Epidemiology, Global Burden, and Pathophysiology

T2D is one of the most prevalent and costly chronic diseases worldwide. It is estimated that more than 530 million adults live with T2D globally, accounting for nearly 10 percent of the adult population.[31] The International Diabetes Federation projects this number to rise to over 700 million by 2045, with the greatest increases expected in low- and middle-income countries undergoing rapid urbanization.[31,32] In the United States, approximately 38 million adults have T2D, of whom about 90 to 95 percent have T2D.[33–35] An additional ~100 million adults are estimated to have prediabetes, reflecting a large population at high risk for disease progression.[36] Despite major advances in prevention and therapy, T2D remains a leading cause of death and disability worldwide.

The global burden of T2D extends beyond its high prevalence. It contributes substantially to cardiovascular disease, chronic kidney disease, visual impairment, and lower-limb amputation. T2D is also a leading cause of premature mortality, responsible for more than 6 million deaths annually. The economic consequences are significant, with global healthcare expenditures of approximately 1 trillion USD each year.[37] The rise in T2D incidence parallels increases in obesity, physical inactivity, and unhealthy diets, yet social, environmental, and genetic determinants also contribute. The burden is unequally distributed, with disadvantaged populations experiencing higher incidence and poorer outcomes due to limited access to preventive care and treatment.[38]

The pathophysiology of T2D is characterized by a progressive decline in insulin sensitivity and pancreatic beta-cell function. In the early stages, insulin resistance develops in skeletal muscle, adipose tissue, and liver, leading to impaired glucose uptake and increased hepatic glucose production.[39] The pancreas compensates by producing more insulin, but this compensatory phase eventually fails due to beta-cell stress, mitochondrial dysfunction, and apoptosis.[40] Chronic hyperglycemia, oxidative stress, and lipotoxicity further exacerbate cellular

injury, creating a cycle that accelerates disease progression. These metabolic abnormalities lead to chronic inflammation, endothelial dysfunction, and atherosclerosis, which underlie many of the vascular complications of T2D.[41]

In addition to genetic predisposition and lifestyle factors, emerging evidence indicates that environmental exposures influence the onset and course of T2D.[42,43] Pollutants such as fine particulate matter, nitrogen dioxide, and airborne toxics can induce systemic inflammation and oxidative stress, processes that are central to insulin resistance and beta-cell failure.[42] These mechanisms align with the broader concept of the exposome, which encompasses the cumulative effects of environmental, behavioral, and social factors across the life course. Understanding how these external exposures contribute to T2D pathophysiology is essential for developing comprehensive prevention strategies that extend beyond traditional behavioral interventions.[44,45]

In California, the burden of T2D mirrors the national and global trends but is amplified by the state's demographic, socioeconomic, and environmental diversity. More than 3.2 million adults in California have diagnosed T2D, and an estimated 5.9 million have prediabetes.[46–48] The prevalence is highest in the Central Valley and parts of Southern California, particularly the Inland Empire (Riverside and San Bernardino counties), Imperial County, and inland portions of Los Angeles County.[47] The economic cost to the state is profound, exceeding 45 billion USD annually in direct and indirect expenses.[46] As one of the most polluted states in the nation, California offers a critical context for studying how environmental and social determinants intersect with metabolic disease risk.

Overview of Air Pollution, Pollutants, and Exposure Assessment

Air pollution is a complex mixture of gases and particles emitted from both natural and anthropogenic sources. Major anthropogenic contributors include fossil-fuel combustion from transportation, industry, and power generation, while natural sources such as wildfires, windblown dust, and sea spray can also contribute substantially to regional air quality. The United States Environmental Protection Agency (EPA) classifies six "criteria" pollutants—particulate matter (PM_{10} and $PM_{2.5}$), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), carbon monoxide (CO), and lead (Pb), which are regulated under the National Ambient Air Quality Standards (NAAQS). Of these, fine particulate matter ($PM_{2.5}$) and O_3 are most frequently linked to adverse health outcomes. $PM_{2.5}$ consists of particles smaller than 2.5 micrometers in diameter that can penetrate deeply into the respiratory tract and enter the systemic circulation, while ozone is a highly reactive gas formed through photochemical reactions between nitrogen oxides and volatile organic compounds under sunlight.[49]

The composition of air pollution varies by region, season, and source type. In urban settings, traffic-related emissions are dominant, producing a mixture of elemental carbon, organic carbon,

nitrates, sulfates, and trace metals. Industrial sources and power plants contribute sulfur compounds and secondary particles formed from gaseous precursors such as sulfur dioxide and nitrogen oxides. In California, geographic features such as mountain basins and temperature inversions exacerbate pollutant accumulation, particularly in the Central Valley and Southern California. These regions often experience some of the highest annual $PM_{2.5}$ concentrations in the United States. Wildfire smoke has also emerged as an increasingly important contributor, adding substantial episodic increases in $PM_{2.5}$ and toxic organic compounds. Seasonal patterns are evident, with wintertime PM dominated by combustion sources and summertime pollution driven largely by photochemical formation of ozone and secondary aerosols.

In addition to criteria pollutants, air contains numerous hazardous air pollutants (HAPs) or "air toxics," including volatile organic compounds, polycyclic aromatic hydrocarbons (PAHs), and metals such as nickel, chromium, cadmium, and lead. These compounds are emitted from industrial processes, fuel combustion, and chemical manufacturing. Many air toxics are carcinogenic or neurotoxic and can cause endocrine and metabolic disturbances even at low concentrations. The Air Toxics Hot Spots Program in California has identified toxic compounds of concern, with benzene, 1,3-butadiene, formaldehyde, and diesel particulate matter representing major contributors to health risk.[50] Recent research also highlights the role of ultrafine particles, microplastics, and secondary organic aerosols as emerging pollutants of potential concern.[51–53] These pollutants are not currently regulated by federal air-quality standards but may contribute to chronic systemic inflammation and oxidative stress, underscoring the need for continued monitoring and research.

Air pollution exposure is assessed using a combination of monitoring networks, satellite observations, and atmospheric modeling.[54] Ground-based monitoring stations operated by the U.S. Environmental Protection Agency and state agencies provide direct measurements of key pollutants such as $PM_{2.5}$, NO_2, and O_3 at high temporal resolution. To fill spatial gaps between monitors, chemical transport models (such as CMAQ and GEOS-Chem) and land-use regression models incorporate meteorology, emissions inventories, traffic data, and topography to estimate concentrations across broader geographic areas.[54] Satellite-derived aerosol optical depth and trace gas retrievals further enhance spatial coverage, especially in regions lacking ground monitors.[55] Increasingly, hybrid models that integrate monitoring, satellite, and modeled data are used to produce high-resolution (1 km or finer) daily estimates suitable for epidemiologic analyses. These methods enable linkage of environmental exposures with health outcomes while accounting for spatial and temporal variability in pollutant distributions.

Emerging linkage between air pollution and T2D

Over the past two decades, a growing body of evidence has revealed that air pollution is not only a respiratory or cardiovascular hazard but also a metabolic risk factor associated with T2D.[42,43] Epidemiologic studies conducted across North America, Europe, and Asia consistently

show higher T2D incidence and prevalence in populations exposed to elevated levels of $PM_{2.5}$,[56] NO_2,[57] and traffic-related pollutants. Early ecological studies first demonstrated a geographic correlation between $PM_{2.5}$ concentrations and T2D prevalence, which was later confirmed in longitudinal cohort analyses with individual-level exposure data.[43] These findings have been replicated in diverse populations and persisted after adjustment for body mass index, physical activity, and socioeconomic status, suggesting an independent contribution of air pollution to metabolic disease risk.

Mechanistic and clinical studies provide biological plausibility for this association. Chronic exposure to air pollutants induces systemic oxidative stress, low-grade inflammation, and endothelial injury, which disrupt insulin signaling and glucose regulation.[42,43,49,58] Inhaled fine particles can promote the release of inflammatory cytokines and reactive oxygen species that impair pancreatic beta-cell function and increase insulin resistance in peripheral tissues.[59] Air pollution has also been linked to altered lipid metabolism, mitochondrial dysfunction, and activation of the hypothalamic–pituitary–adrenal axis, all of which contribute to impaired glucose homeostasis.[60] These mechanisms parallel those implicated in obesity-related metabolic dysfunction, suggesting that air pollution acts synergistically with traditional risk factors to accelerate the development of T2D.

The emerging recognition of air pollution as a metabolic stressor has significant public health implications. The global prevalence of T2D continues to rise despite advances in lifestyle and pharmacologic prevention, indicating that environmental exposures may undermine traditional control strategies. Populations residing in urban and industrialized areas face continuous exposure to pollutant mixtures that include not only regulated criteria pollutants but also air toxics such as benzene, 1,3-butadiene, and polycyclic aromatic hydrocarbons.[61] These exposures disproportionately affect socioeconomically disadvantaged and racially marginalized communities, amplifying existing health disparities. Understanding the pathways through which air pollution contributes to T2D provides a foundation for developing integrated prevention frameworks that combine clinical, environmental, and policy interventions to reduce the burden of metabolic disease.

Biologic Mechanisms linking air pollution with T2D mellitus

The pathways through which ambient air pollutants influence T2D risk and outcomes are complex and involve a cascade of local, systemic, and organ-specific processes (Figure 1).[58] Exposure to $PM_{2.5}$, PM_{10}, NO_2, and O_3 initiates biological responses that impair insulin sensitivity, disrupt glucose metabolism, and promote the development and progression of T2D and its complications.[42,49]

Air pollution exposure first triggers oxidative stress and inflammatory signaling within the respiratory tract. Reactive oxygen species and reactive nitrogen species are generated by pollutant particles and gases, activating toll-like receptors and other sensing pathways. This initiates the release of inflammatory cytokines, chemokines, and acute phase reactants that spread beyond the lungs and enter systemic circulation. Endothelial barrier disruption and neurohumoral activation, including stimulation of the hypothalamic–pituitary–adrenal axis and sympathetic nervous system imbalance, further amplify these early inflammatory responses.

The biological signals initiated in the lungs are then transmitted throughout the body. Circulating oxidized lipids, acylcarnitines, microparticles, and cytokines contribute to insulin resistance in peripheral tissues. Experimental studies have demonstrated that exposure to concentrated particulate matter increases plasma free fatty acids and inflammatory markers,[62] while antioxidant interventions can mitigate these effects.[63] Pollutants may also translocate directly into the bloodstream, adding to the systemic oxidative and inflammatory burden.[64]

Downstream effects occur in several insulin-responsive organs. In adipose tissue, chronic exposure to air pollution recruits pro-inflammatory monocytes and macrophages, shifting immune balance toward a Th1/M1 phenotype that promotes inflammation, impaired lipolysis, and insulin resistance. In the liver, exposure induces steatohepatitis-like changes with altered lipid metabolism, endoplasmic reticulum stress, and disrupted insulin signaling.[11] In skeletal muscle, endothelial dysfunction and reduced glucose transporter (GLUT4) activity impair insulin-stimulated glucose uptake.[65] In the central nervous system, hypothalamic inflammation, blood–brain barrier disruption, and altered autonomic regulation contribute to impaired energy balance, elevated cortisol, and disrupted circadian rhythms.[66,67]

Additional mechanisms have also been identified. Pollution can reduce thermogenic activity in brown adipose tissue, lowering energy expenditure through mitochondrial and transcriptional dysregulation.[67] Chronic particulate matter exposure promotes chromatin remodeling and methylation changes in genes linked to inflammation and insulin resistance, effects that may be partially reversible after exposure reduction.[67] Circadian rhythm disruption has emerged as another common feature of pollutant exposure, contributing to the development of metabolic syndrome. Neurohormonal activation, including heightened sympathetic activity and hypothalamic–pituitary–adrenal stimulation, raises blood pressure and catecholamine levels, worsening insulin resistance and accelerating T2D progression.[68]

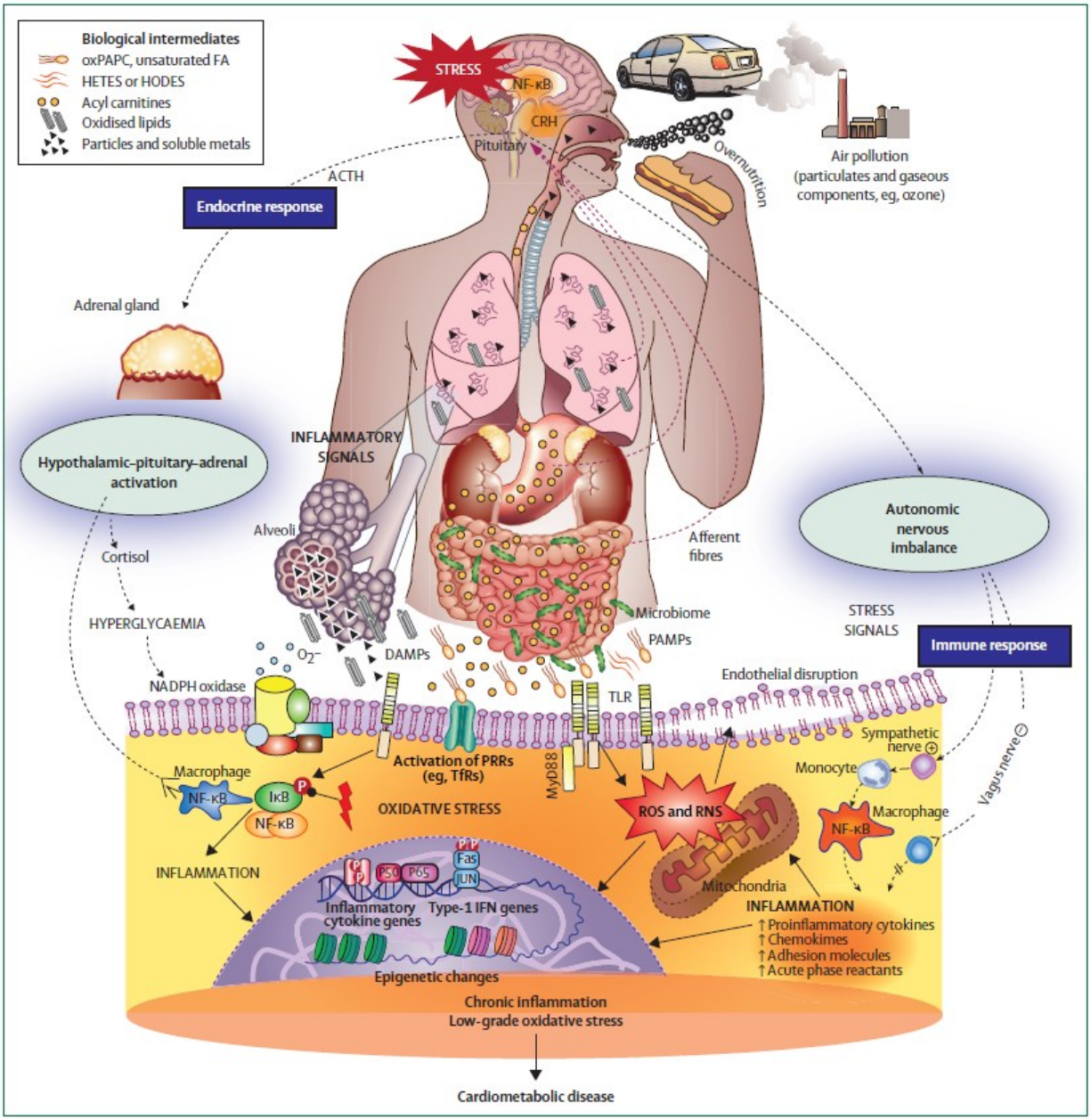


Figure 1. Mechanisms of $PM_{2.5}$-mediated metabolic and cardiovascular effects (obtained from Rajagopalan et al. Lancet Diab Endoc 2024)

Review studies linking long-term exposure to air pollution and T2D

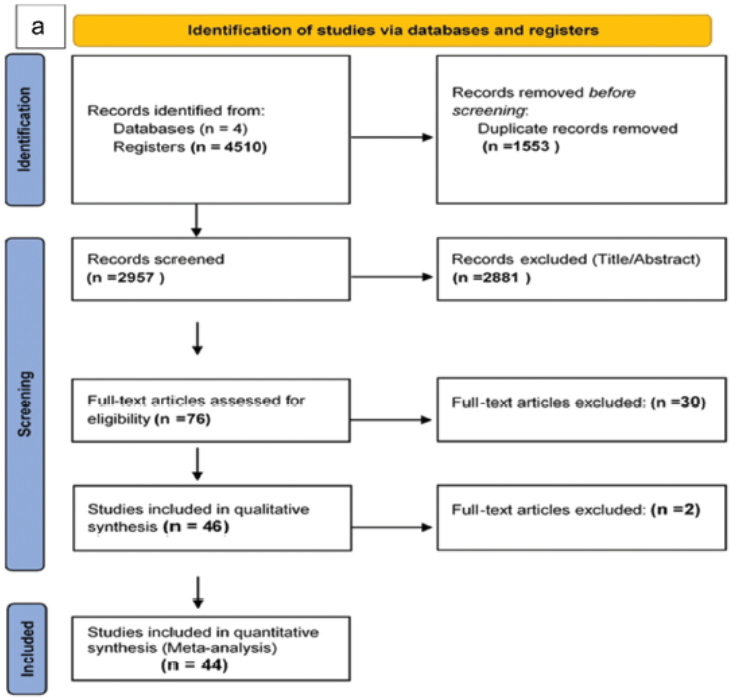
Long-term exposure to ambient air pollutants such as $PM_{2.5}$, NO_2, and O_3 has been increasingly linked to the development of T2D. Large prospective cohorts and administrative database studies across North America, Europe, and Asia have shown consistent associations between chronic $PM_{2.5}$ and NO_2 exposure and higher incidence of T2D. These pollutants may impair glucose homeostasis and insulin sensitivity through systemic inflammation, oxidative stress, and endothelial dysfunction. Long-term exposure can also alter adipose tissue metabolism,

promote low-grade chronic inflammation, and accelerate vascular and metabolic aging, all of which are central to T2D pathophysiology.

Evidence is strongest for $PM_{2.5}$, where concentration–response relationships have been observed even at levels below current regulatory standards. Although findings for O_3 are more variable, some studies suggest it may contribute to impaired glucose regulation through oxidative mechanisms, particularly in warmer seasons or in combination with co-pollutants. Importantly, these associations remain significant after adjusting for individual and neighborhood-level socioeconomic factors, suggesting an independent role of air pollution in T2D risk. Overall, the growing epidemiologic evidence highlights long-term exposure to air pollutants as a modifiable environmental determinant of T2D, underscoring the need for stringent air quality standards and targeted prevention efforts.

Several systematic reviews and meta analyses have examined the associations between long-term exposure to air pollutants and T2D, generally showing similar directionality of associations and similar effect sizes.[7,69–75]

Azizi et al. [76] conducted a global systematic review and meta-analysis to evaluate the association between ambient $PM_{2.5}$ and PM_{10} and T2D. A comprehensive literature search was performed on November 4, 2022, using four major databases: PubMed, Embase, Web of Science, and Scopus. The search strategy was designed through a structured, multi-step process and restricted to English-language, human studies reporting original epidemiologic research. Reviews, animal studies, clinical trials, and studies of indoor air pollution were excluded. Eligible studies were required to examine outdoor $PM_{2.5}$ or PM_{10} exposure in relation to T2D prevalence or incidence, or related glycemic endpoints, and to report effect estimates such as odds ratios or risk ratios. A wide range of observational study designs was eligible, including cohort, case-control, cross-sectional, panel, time-series, and case-crossover studies. Across all databases, 4,510 records were initially identified. After removal of 1,553 duplicates, 2,957 titles and abstracts were screened,



leading to exclusion of 2,881 non-relevant studies. Seventy-six full-text articles were assessed for eligibility, all of which evaluated $PM_{2.5}$ or PM_{10} in relation to T2D outcomes and were included in the qualitative synthesis. Of these, 46 studies met predefined quality criteria and were retained for quantitative meta-analysis. The included studies spanned 2010 to 2022, with more than half published after 2018, reflecting rapid recent growth in the literature. Research was conducted across 18 countries, with substantial geographic diversity. Metaanalysis showed that each 10 μg/m³ increase in $PM_{2.5}$ was associated with a 9% higher odds of prevalent T2D (OR = 1.09, 95% CI: 1.05–1.14) and a 9% higher risk of incident T2D (HR = 1.09, 95% CI: 1.02–1.17). For PM_{10}, the odds of prevalent T2D were increased by 13% per 10 μg/m³ (OR = 1.13, 95% CI: 1.06–1.19), while the risk of incident T2D rose by 10–24% depending on study inclusion (HR = 1.10–1.24). Subgroup analyses suggested stronger associations in North America for $PM_{2.5}$ and in Asia for PM_{10} (Figure 2). Mechanistic evidence supports biological plausibility through systemic inflammation, oxidative stress, and endoplasmic reticulum stress, which together impair insulin signaling, damage pancreatic β-cells, and promote insulin resistance. Other pollutants are less well studied. A recent systematic review and meta-analysis of five studies examining long-term O_3 exposure and T2D reported that each 10 μg/m³ increase in ambient O_3 was associated with a 6% higher risk of T2D (pooled effect estimate 1.06; 95% CI 1.02–1.11).[77] In addition, a 2014 meta-analysis of prospective and cross-sectional studies found that each 10 μg/m³ increase in long-term exposure to NO_2 was associated with 13% increase in hazards of incident T2D (pooled adjusted hazard ratio 1.13; 95% CI 1.01–1.22) across prospective cohorts, with moderate heterogeneity and consistent adjustment for major confounders including age, sex, body mass index, and smoking.[78] It is important to note that many of the studies examining long-term exposures and incident T2D were conducted over a decade ago, preceding major shifts in emission sources, regulatory standards, and wildfire frequency.

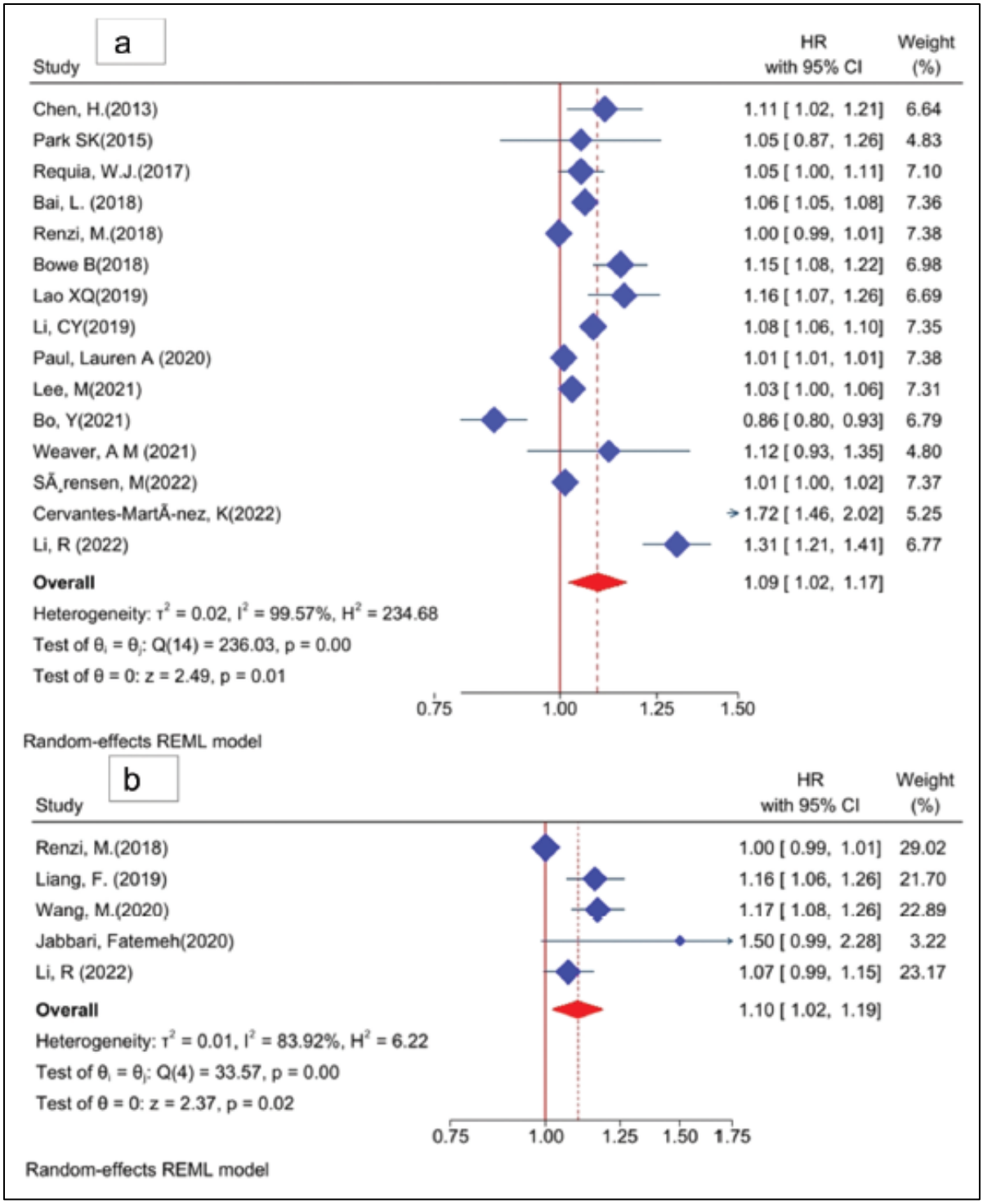
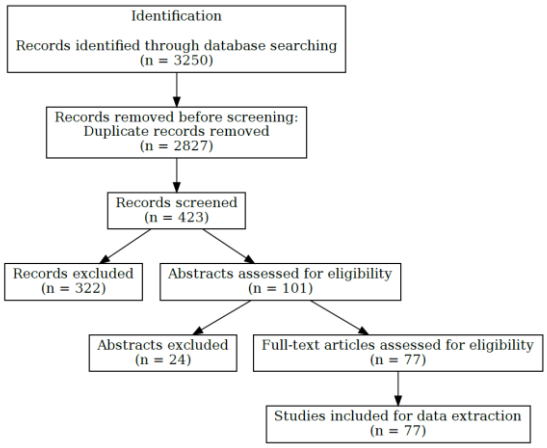


Figure 2. Association between $PM_{2.5}$ (a) and PM_{10} (b) exposure and T2D incidence. (Azizi et al. 2025)

Studies linking short-term exposure to air pollution and T2D

Short-term fluctuations in ambient air pollution can acutely influence glycemic control, metabolic stability, and the risk of diabetes-related complications. Acute increases in $PM_{2.5}$, PM_{10}, NO_2, and O_3 have been associated with short-term metabolic deterioration and increased healthcare utilization among individuals with or at risk for diabetes. These pollutants can provoke systemic inflammation, oxidative stress, endothelial dysfunction, and sympathetic activation within hours of exposure. Such physiologic responses can raise circulating glucose, increase insulin resistance, and impair vascular tone, triggering metabolic decompensation in vulnerable populations. In those with pre-existing diabetes, these mechanisms may precipitate hyperglycemic crises, cardiovascular events, infections, and other acute complications that require emergency care.

We conducted a systematic review of short-term air pollution exposure and type 2 diabetes–related outcomes using a comprehensive search of PubMed, Embase, and MEDLINE for human studies published in English between January 2000 and June 2025. The search identified 3,250 records, of which 423 were screened by title, 101 underwent abstract review, and 77 met inclusion criteria after full screening. The included studies spanned 22 countries, with the largest contributions from China, Canada, Italy, and the United States, providing broad geographic representation.



The evidence base is dominated by time-series studies (approximately 30) and time-stratified case-crossover designs (approximately 25), with smaller numbers of cohort or panel studies, cross-sectional or ecological analyses, and other specialized approaches. Most studies evaluated acute exposure windows from same-day through lag 3 to 5 days, with some extending to lag 7 to 14 days. Common analytic methods included generalized additive models, Poisson or quasi-Poisson regression, conditional logistic regression, and distributed lag nonlinear models, allowing assessment of both immediate and delayed effects.

$PM_{2.5}$ was by far the most frequently studied pollutant, followed by PM_{10}, NO_2, and O_3, often in multi-pollutant models. Several studies also examined PM components, wildfire or dust-related events, composite air quality indices, and co-exposures such as temperature or traffic noise. Exposures were most commonly expressed as per 10 $\mu g/m^3$ increases, daily mean concentrations, interquartile range increments, or short-term moving averages.

Primary outcomes focused on mortality, hospitalizations, and emergency department visits, particularly for diabetes, cardiovascular disease, and respiratory conditions. Secondary analyses frequently explored effect modification by age, sex, comorbidities, season, and socioeconomic context, with fewer studies examining biomarkers or metabolomic outcomes. Overall, the short-term literature consistently relies on robust time-series and case-crossover designs to demonstrate that acute increases in air pollution are associated with higher risks of diabetes-related morbidity and mortality, while also highlighting heterogeneity by pollutant, outcome, and population subgroup.

Across pollutants, $PM_{2.5}$ was the most frequently studied in relation to diabetes-related hospitalizations[79–85] (including a study on length of stay and hospitalization cost[86]), physician visits,[87] ambulance dispatch/paramedic assessment,[88] and mortality.[89,90] Most studies evaluated $PM_{2.5}$ concentration, but some studies evaluated wildfire specific $PM_{2.5}$.[88,91] Generally most,[79–81,83,85,87,89,91] but not all,[92–94] studies have shown positive relationships between $PM_{2.5}$ and diabetes outcomes. Typical findings indicated that each 10 μg/m³ (or interquartile range/standard deviation) increase in $PM_{2.5}$ corresponded to a 0.5% to 3% rise in acute healthcare encounters or mortality.[79–81,84,89,90] For $PM_{2.5}$ and diabetes-related outcomes, lag 0–2 days[79,82,83,88–90] is the most common and significant window analyzed across studies, however, some studies extended it longer to 3 days, 5 days,[87] 7 days[92] and up to 16 days (moving average exposure).[86] In studies comparing different lags, generally shorter lags (0-2 days) corresponded to stronger effect size,[79,80,85,89,90] however, some studies showed longer lags have stronger effects (lag 6-8 days[81,87]). Heterogeneity was examined in some studies: by temperature,[79] season,[87,92,93] age,[90,92,94] gender,[90,92] comorbidities,[89] regional variation,[89,94] wildfire vs non-wildfire $PM_{2.5}$.[91] The results were heterogenous overall. In one study, $PM_{2.5}$ effects were noted to be strongest at low and moderate temperatures, indicating climate-modified heterogeneity.[79] In another study, stronger associations were noted in cold seasons and among younger and male subgroups.[87]

Other studies have examined the association between PM_{10} with diabetes outcomes: hospitalizations,[81,84,90,92,93,93,95] hospitalization characteristics including cost/length of stay,[86] emergency department visits,[96] and mortality.,[94,97,98] often examined together with $PM_{2.5}$. Overall, most,[84,85,90,92,95,97,98] but not all,[86,93] studies have shown positive relationship between PM_{10} and diabetes outcomes. short-term PM_{10} exposure associated with increased diabetes-related hospitalizations, emergency visits, and mortality, typically in the range of 0.3%–3% per 10 μg/m³ increase.

Fewer studies examined SO_2 and diabetes outcomes.[82,84,86,90,92,93,96,98] Some studies have shown positive relationship with adverse events (mortality[98]/hospitalizations[82,84]/length of stay[86]) while others did not.[92,93,96] Among the positive studies, An interquartile range or 10 μg/m³ rise in SO_2 concentration corresponded to roughly 0.5% to 3.8% higher T2D-related hospitalizations or deaths, with the most pronounced effects observed within 0–3 days of exposure.[82,84,90,98] In one

study, the association between SO_2 and T2D hospitalizations remained significant after adjustment for O_3 but became null when $PM_{2.5}$ was included in the model. Similarly, another study found that the SO_2–T2D mortality relationship was independent of NO_2, yet adjustment for either $PM_{2.5}$ or PM_{10} eliminated the association.[90] Lag-specific analyses further indicated that the effect of SO_2 on T2D hospitalizations peaked at a lag of 3 days in one study[84] and at a lag of 1 day on T2D mortality in another.[90]

For NO_2, studies have examined outpatient visits,[87,99] ED visits,[99] hospitalizations,[82,84,86,92,93,96,99] and mortality.[89,90,94,98]. Across 12 studies, short-term exposure to nitrogen dioxide was consistently linked with adverse diabetes-related outcomes, including increased hospitalizations, outpatient visits, and mortality. For example, In Lanzhou, China, Ye et al. reported a 3.4% increase in T2D outpatient visits per 10 μg/m³ rise in NO_2 (lag 0-3 days)[87]; Yin et al.[100] observed a 3.96% (per IQR of NO_2) increase in diabetes mortality at lag 0–2 days across all administrative regions in China;[89] Zhang et al. found a 2.2% increase in diabetes hospitalizations per 10 ug/m3 at lag 0-4 days;[92] and Gariazzo et al. reported a 7.3% increase metabolic mortality per 10 μg/m³ of NO_2 in Italy (lag 0-5 days).[94] Most studies employed time-series or case-crossover designs with lags typically spanning 0–7 days. Generally, shorter lag was associated with stronger effect.[82] The observed relationship was linear or near-linear.[82,89] Associations were generally independent of other pollutants such as $PM_{2.5}$, PM_{10}, SO_2, CO, and O_3.[82] Younger individuals,[87] and those exposed during colder seasons[92] were often more vulnerable. Collectively, these findings underscore NO_2 as a significant contributor to short-term diabetes-related health burdens across diverse geographic regions.

The studies on O_3 exposure examined a range of short-term diabetes-related outcomes, including hospitalizations for T2D or its complications,[82,83,92,93,96,99] outpatient visits for diabetes,[87,99] emergency department,[99] and mortality from diabetes and related complications.[89] Together, these studies assessed both acute healthcare utilization and fatal outcomes as indicators of short-term metabolic stress and diabetes exacerbation associated with ambient ozone exposure. Across the studies examining O_3 and diabetes outcomes, evidence points to heterogeneous and non-consistent associations. Some studies showed positive relationships between O_3 and diabetes outcomes. In a study in Lanzhou, China, Ye et al. showed positive relationship between O_3 and outpatient visits for diabetes (per 10 ug/m3 increase in maximum of 8h averaged O3 in a day ($O_3$8h) at lag05 (RR 1.012, 95% CI: 1.001, 1.023).[87] In another study in all administrative regions in China, each 47.3 ug/m3 of O_3 (maximum effect at lag 0-2 days) was associated 2.15% increase in diabetes mortality, but this was only observed at concentrations exceeding 60 ug/m^3.[89] However, large studies in Canada,[99] China,[82,83,92] South Korea[96] and Bulgaria[93] showed no associations between O_3 and adverse diabetes outcomes. Inconsistencies across studies may stem from differences in exposure metrics, population characteristics, and regional factors. Ozone levels vary by season and meteorology, and its effects may appear only

above certain thresholds. Co-pollutant confounding, differences in healthcare access, and variation in study design further contribute to the mixed findings.

Air Toxics and T2D

Human epidemiologic studies consistently demonstrate associations between exposure to air toxics, particularly polycyclic aromatic hydrocarbons (PAHs), volatile organic compounds (VOCs) such as benzene and 1,3-butadiene, and persistent organic pollutants including dioxins and dioxin-like compounds, and impaired glucose homeostasis, insulin resistance, prediabetes, and T2D. Much of this evidence derives from population-based studies using biomarker-based exposure assessment, including urinary or serum metabolites measured in large surveys such as NHANES, as well as occupational and environmentally exposed cohorts.[101–103]

Across multiple studies, higher internal doses of PAHs are associated with increased odds of diabetes and insulin resistance, with consistent findings for metabolites of naphthalene, fluorene, phenanthrene, and pyrene.[101,104,105] These associations are observed in the general population, occupational cohorts such as coke oven workers, and meta-analyses, and are often dose dependent.[103,105] Subgroup analyses frequently suggest stronger associations among women, younger adults, nonsmokers, and individuals without obesity, indicating susceptibility beyond traditional metabolic risk factors.[102,106]

Dioxins and dioxin-like compounds, including TCDD and dioxin-like PCBs, have been linked to diabetes prevalence and incidence in environmentally exposed populations and occupational cohorts. Studies of residents near industrial contamination or waste incineration sites, as well as military cohorts exposed during the Vietnam War, report higher diabetes risk with increasing serum dioxin burden.[107–109] Meta-analytic evidence indicates elevated risk in both sexes, with potential modification by exposure intensity and exposure mode.[110] These epidemiologic findings are supported by mechanistic reviews highlighting aryl hydrocarbon receptor–mediated disruption of metabolic regulation.[111,112]

Emerging population-level evidence also implicates volatile organic compounds, including benzene and 1,3-butadiene, in dysregulation of glucose metabolism. Recent analyses of U.S. population data demonstrate associations between urinary metabolites of 1,3-butadiene and indices of glucose homeostasis, prediabetes, and diabetes, with mediation by inflammatory and hepatic pathways such as alkaline phosphatase.[113] Exposome-wide association studies further suggest that mixtures of VOCs contribute to insulin resistance risk.[114]

Mechanistically, these human associations are biologically plausible. Air toxics are linked to insulin resistance, beta-cell dysfunction, oxidative stress, and chronic low-grade inflammation,

often involving activation of the aryl hydrocarbon receptor, suppression of PPAR signaling, endocrine disruption, and perturbation of lipid and glucose metabolism.[111,112,114]

Taken together, the human literature provides consistent evidence that exposure to multiple air toxics is associated with increased risk of T2D. These findings extend beyond criteria air pollutants and suggest that combustion-related toxicants and persistent organic pollutants contribute meaningfully to the global diabetes burden, operating through metabolic and inflammatory pathways that complement traditional cardiometabolic risk factors.

Gaps in knowledge

Although global evidence links long-term air pollution exposure to T2D, contemporary data from the United States remain limited, especially in regions characterized by complex pollutant mixtures and wildfire events. Many landmark studies were conducted over a decade ago, preceding major shifts in emission sources, regulatory standards, and wildfire frequency. As a result, there is a scarcity of recent data evaluating how modern pollutant profiles—particularly fine particulate matter enriched with combustion byproducts—affect diabetes risk in the current U.S. context. California, where wildfire smoke now represents a dominant and episodically extreme exposure source, lacks comprehensive epidemiologic studies quantifying short-term metabolic impacts. The evolving composition of $PM_{2.5}$ and co-pollutants from wildfire events may have distinct biological effects, yet few clinical studies have examined this within contemporary diabetic populations.

Another major limitation is the lack of research in racially, ethnically, and socioeconomically diverse U.S. populations, particularly those most affected by both diabetes and environmental burdens. Many existing datasets are derived from homogenous cohorts in Europe or older administrative data that inadequately capture the demographic and environmental complexity of California. Moreover, while evidence for chronic exposure and diabetes incidence is robust, the level of evidence linking short-term pollutant fluctuations to acute diabetes outcomes remains relatively weak and inconsistent. Few studies have integrated high-resolution exposure modeling with administrative health data to evaluate temporal lags, dose–response relationships, and modifiers such as temperature or wildfire smoke. Addressing these gaps through a California-based study using refined estimates of criteria pollutants and real-world diabetes outcomes would fill critical voids in the current literature and provide insights relevant to contemporary environmental and public health challenges

Task 2. Develop daily air pollution models and surfaces for criteria pollutants

For Task 2, the development of daily high-resolution air pollution models and exposure surfaces builds on a series of related CARB-funded projects, including 21RD004, 22RD010 (this project) and 23RD011, which collectively contributed to the development of statewide criteria

pollutant modeling across different time periods. This contract represents the first application of these harmonized daily exposure surfaces to a comprehensive set of diabetes-related health endpoints across multiple statewide datasets.

Methodology

Acquiring and processing air pollution data from regulatory monitoring

We acquired and processed daily air pollution data and their spatial locations from the U.S. Environmental Protection Agency (https://aqs.epa.gov/aqsweb/airdata/download_files.html). The regulatory data measurements were obtained from monitoring sites equipped with standardized instruments for measuring air pollutants. Specifically, NO_2 was measured using instruments coded as 42602, which typically involve chemiluminescence techniques, recognized for their accuracy in detecting nitrogen dioxide levels in ambient air. $PM_{2.5}$ concentrations were measured using Federal Reference Method (FRM) or Federal Equivalent Method (FEM) instruments coded as 88101, which involve either gravimetric or continuous monitoring techniques to capture fine particulate matter in the air. Ozone (O_3) measurements were conducted using instruments coded as 44201, which commonly utilize ultraviolet photometry to accurately measure ozone concentrations. In California, the spatial distribution of the regulatory air quality monitoring data for NO_2, $PM_{2.5}$ and O_3 are presented in Figure 3 (left for NO_2, middle for $PM_{2.5}$ and right for O_3) and the respective unique number of regulatory sites is presented in Table 1. The number of criteria pollutant monitoring sites exhibits overall year-to-year stability, reflecting the long-term design of CARB's monitoring network. Most sites operate continuously for extended periods to support trend analysis and regulatory assessment. However, modest fluctuations in site counts do occur over time due to programmatic considerations such as instrument maintenance or replacement, temporary suspensions for quality assurance, site relocations, or targeted additions/removals to address emerging regulatory or scientific priorities. These adjustments are made by CARB as part of routine network management and do not indicate substantial changes in monitoring strategy. Consequently, while the core monitoring network remains stable, small year-to-year variations in the number of available sites are expected and are reflected in the data used in this study.

The trend for NO_2 measurement sites shows a slight decline during the early 1990s, with the number of unique sites decreasing from 151 in 1990 to 147 in 2000. This downward trend continued until 2006, when the number of monitoring sites reached its lowest point. After 2006, the number of unique NO_2 measurement sites fluctuated between 127 and 135, suggesting variability in monitoring efforts. Overall, there is no consistent upward or downward trend in NO_2 monitoring, indicating that the focus on this pollutant has varied over the years. The total number of unique NO_2 air quality monitors is 277. In contrast, the trend for $PM_{2.5}$ reveals a clear upward trajectory in the number of unique measurement sites. Starting with 183 sites in 1999, the number steadily increased to 252 by 2021. This growth is particularly evident from 2000

onward, demonstrating a growing recognition of the importance of this pollutant and dedicated resources to understanding and mitigating its impacts. The total number of unique $PM_{2.5}$ air quality monitors is 331. For O_3, the trend indicates a generally stable pattern with a gradual increase in monitoring sites over time. The number of unique O_3 measurement sites increased from 194 in 1990 to 198 in 2008, with some fluctuations throughout the years. Although the overall growth in O_3 monitoring efforts is less pronounced than that of $PM_{2.5}$, it still demonstrates a steady commitment to tracking this pollutant. The total number of unique O_3 air quality monitors is 379.

In our modeling process, we also applied fixed site saturation monitoring data in our analysis. A detailed description of the saturation monitoring data can be found in Supplementary File 3 and in a previously published paper.[115]

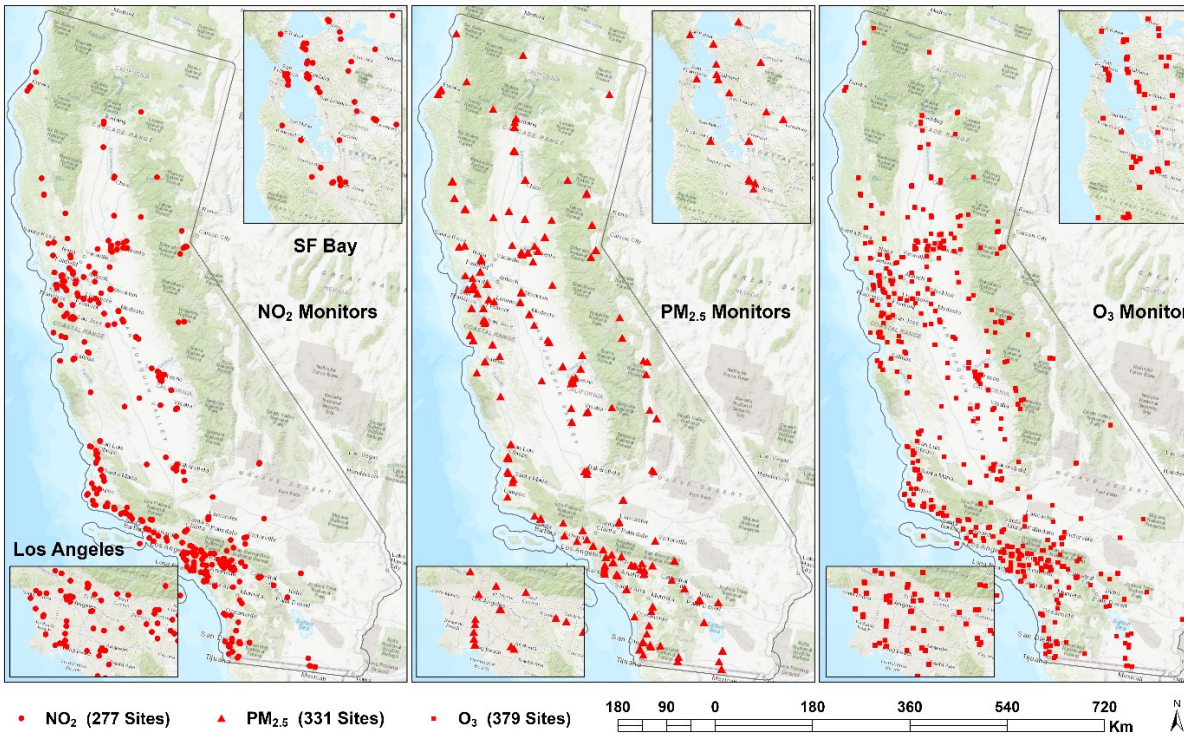


Figure 3. The spatial distributions of the regulatory monitors for NO_2 (left panel), $PM_{2.5}$ (middle panel), and O_3 (right panel) over the observable time periods.

Table 1. The unique number of regulatory monitoring sites with the respective effective measurements of NO_2, $PM_{2.5}$ and O_3 across the study period.

Year	**NO2**	**PM2.5**	**O3**
1989			182
1990	151		194
1991	150		201
1992	149		205
1993	159		199
1994	164		208
1995	163		
1996	159		
1997	156		
1998	154		
1999	148	183	
2000	147		
2001	153		
2006	127		186
2007	129	213	195
2008	136	221	198
2009	130	225	192
2010	132	228	194
2011	127	229	196
2012	132	248	198
2013	129	242	190
2014	132	246	189
2015	133	240	185
2016	135	238	185
2017	132	240	184
2018	129	246	180
2019	128	241	181
2020	124	247	182
2021	127	252	178
Total[ξ]	277	331	379

[ξ]: Total number of unique air quality monitoring sites across the study period (1989–2021).

Acquiring and processing air pollution data from Google Streetcar monitoring

Google Streetcar had mobile monitoring of the three criteria pollutants across San Francisco Bay (counties of Alameda, Contra Costa, San Francisco and San Mateo), Los Angeles County, and Central Valley regions (see: https://www.google.com/earth/outreach/special-projects/air-quality/). The Google Streetcar mobile measurements for each region are highly spatially autocorrelated due to the intense sampling of air pollutants on its road network. To ensure that our models captured a wide range of variability in road traffic patterns while minimizing the influence of spatial autocorrelation, we selected 150 road segments for each region through a location-allocation algorithm.[116] The location-allocation algorithm is deterministic and was intentionally used to ensure spatially representative coverage of traffic conditions within each region without spatial autocorrelation rather than to support inference at the individual road-segment level. Because the objective was to characterize regional traffic patterns for exposure modeling, rather than to evaluate the effect of specific segments, we were not interested in percentage of mobile sampling being used and did not conduct sensitivity analyses based on alternative segment selections. Spatial autocorrelation can lead to inflated model performance metrics and reduced generalizability by over-representing certain areas or patterns. By using the location-allocation algorithm, we distributed the selected road segments more evenly across each region, reducing clustering and ensuring that our models are better representative of the broader spatial patterns across California. This approach helped in developing more robust and interpretable models by preventing overfitting localized traffic conditions. A total of 150 road segments with each road segment having at least 100 measurements was selected for each of the four regions: Alameda and Contra Costa; San Francisco and San Mateo; Los Angeles, and Central Valley. Each region had (1) 50 road segments selected from locations within 500 m of highways allowing truck traffic, or within 500 m of major California ports (i.e., goods movement corridors or GMCs), (2) 50 road segments selected from locations within 500 m of highways not allowing truck traffic or within 300 m of major roadways (i.e., non-goods movement corridors or NGMCs), and (3) locations not encompassed in the first and second parts (i.e., control areas or CTRLs). The detailed selection process is documented in Supplementary File 3. A total of 150 segments were selected in each of four regions (600 total), yielding 8,345 daily traffic measurements. These data were used in conjunction with substantially larger datasets from regulatory monitoring (676,612 daily measurements) and saturation monitoring (4,893 daily measurements), ensuring that exposure estimates were not driven by the selected segments alone but reflected broader regional traffic variability.

The Google Streetcar measured NO_2 and O_3 concentrations in the unit of ppb – the same as regulatory monitoring; however, $PM_{2.5}$ concentrations were in total number of particles instead of the typical concentrations in μg m^{-3}. The daily concentration of $PM_{2.5}$ in μg m^{-3} of road segment *i* of traffic corridor *k* on day *j* was estimated through:

$$C_{i,j,k} = G_{i,j,k} * \widehat{R_{J,k}}/\widehat{G_{J,k}} \quad (1)$$

where $C_{i,j,k}$ and $G_{i,j,k}$ represent the converted and original measures. $\widehat{R_{J,k}}$ and $\widehat{G_{J,k}}$ are respectively the mean $PM_{2.5}$ concentrations in μg m^{-3} from all the regulatory monitors and the mean $PM_{2.5}$ particle numbers from all the selected 50 road segments for day j in corridor k. The $PM_{2.5}$ concentrations were estimated separately for each region.

Acquiring and processing air pollution predictors from the observation period

For the predictors (Table 2), the availability of daily traffic data varied across 12 California Department of Transportation (Caltrans) districts, with the earliest traffic data available from 2000 to 2005. We used the data collected by the Caltrans Performance Measurement System (PeMS) to derive roadway daily traffic. PeMS data are collected in real-time from nearly 40,000 individual detectors spanning the freeway system across all major metropolitan areas of the State of California and provide an Archived Data User Service that provides over fifteen years of data for historical analysis. The detector measured traffic flow covered ~5 % highway segments, and we summed hourly traffic to daily traffic for all the stations across California. The interconnected steps were then used to derive daily traffic for all the California highways. Please refer to the Supplementary File 3 for the details of traffic assignment.

The land use data was derived from the statewide parcel data in 2019, combined by the California Air Resources Board (CARB) from individual County Assessor's Offices, and we considered them consistent across all the years. The land cover data was acquired from the National Land Cover Database (NLCD) at five-year intervals (2001, 2006, 2011, 2016, and 2019)[117]. The assumption was that land cover remained constant until the subsequent available measurement. Vegetation dynamics were assessed through the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument-derived data, specifically the Normalized Difference Vegetation Index (NDVI)[118], computed at 16-day intervals since 2000. We assumed the vegetation index remained constant from its previous measurements within 16 days. Daily meteorological data were acquired from the GridMet dataset[119], covering 1989 to 2021 at a 4 km spatial resolution. For satellite remote sensing data, daily measurements from the Ozone Monitoring Instrument (OMI)[120] for NO_2 and O_3 were accessible from 2005 to 2021. The aerosol optical depth (AOD) data[121] was available from 2000 to 2021.

Table 2. LUR predictors and available time periods in the modeling process.

Variables	Source	Spatial Resolution	Temporal Resolution	Time Period	Extension Period
Traffic[δ]	CalTrans	30 m	Daily	2005-2021	1989-2004
Land use[θ]	CARB	30 m	One time	2019	Use 2019
Land cover[¥]	NLCD	30 m	Every 5 years	2001-2019	Use 2001
Vegetation index (NDVI)[€]	MODIS	250 m	Every 16 days	2000-2021	1989-1999
Meteorological data[£]	GridMet	4 km	Daily	1989-2021	None
AOD data[ξ]	MAIAC	1 km	Daily	2000-2021	1989-1999
OMI-NO_2 data[ξ]	NASA's OMI	25 km	Daily	2005-2021	1989-2004
OMI-O_3 data[ξ]	NASA's OMI	25 km	Daily	2005-2021	1989-2004
Distance to highway and major roadways[ǂ]	ESRI	30 m	One time	2018	None
Distance to coast[ǂ]	USGS	30 m	One time	2015	None
Elevation from digital elevation model[ǂ]	USGS	30 m	One time	2015	None
Distance to ports[ǂ]	ESRI	30 m	One time	2018	None

[δ]: Traffic data are derived from the California Department of Transportation (CalTrans)
[θ]: Land use data are provided by the California Air Resources Board (CARB), which combined the parcel data from all the 58 counties in California.
[¥]: Land cover data is derived from the NLCD (National Land Cover Database) provided by the U.S. Geological Survey (USGS).
[€]: The NDVI (Normalized Difference Vegetation Index) data is provided by MODIS (Moderate Resolution Imaging Spectroradiometer) from NASA's Earth Observing System (EOS).
[£]: The meteorological data is sourced from GridMet provided by the University of Idaho.
[ξ]: MAIAC AOD data: Data from the Multi-angle Implementation of Atmospheric Correction (MAIAC) algorithm using MODIS Terra and Aqua satellites; OMI-NO_2 and OMI-NO_3 data are derived from the National Aeronautics and Space Administration Ozone Monitoring Instrument.
[ǂ]: Traditional predictors include distance to the nearest highway and major roadway derived from the ESRI Street data layer for 2018, distance to coast and elevation data derived from the USGS for 2015, and distance to major ports derived from the ESRI data layer for 2018.

Developing daily air pollution models through ML integrated LUR approach

The Deletion/Substitution/Addition (D/S/A) algorithm initiates the selection process by starting with a base model, typically the intercept-only model unless a different starting point is specified. The algorithm then iteratively adds, deletes, or substitutes terms to improve the model's predictive performance. During each iteration, potential modifications to the model, such as adding polynomial terms or interaction effects, are evaluated based on a predefined criterion, usually the reduction of the cross-validated error or the improvement in another model performance metric. The selection process continues iteratively, with the algorithm testing various combinations of terms and retaining the modifications that lead to the greatest improvement in model performance. This process is similar to a guided search through the space of possible models, where each step is evaluated to ensure it moves toward a better fit. The algorithm halts its iterations when no further modifications result in a significant improvement in the model's performance, according to the predefined stopping criteria. These criteria could include a threshold for the minimum improvement in cross-validated R-squared or reaching a maximum number of iterations (15 in our research). At this point, the model with the optimal combination of terms is selected as the final model, representing the best balance between complexity and predictive accuracy. To enhance the interpretability of our modeling results, we limited the predictors to linear terms and avoided interaction terms.

For regulatory and saturation monitoring data, each was treated independently, randomized, and divided into 10 equal folds without considering spatial or temporal constraints. The Google Streetcar data, which spans multiple regions, was randomized and divided into 10 folds separately for each region. These region-specific folds were then merged with the corresponding folds from the other regions, as well as with the 10 randomized folds from the regulatory and saturation monitoring datasets. This approach ensured that each of the 10 folds contained a balanced mix of data from all monitoring types and regions. One subsample was then retained as validation data, while the remaining 9 subsamples served as training data during the modeling process. This cross-validation process was repeated 10 times, with each subsample used once as validation data.

In developing the daily LUR models for NO_2, $PM_{2.5}$, and O_3, we constructed respective models using only available observable data for both predictors and air quality measures. Collinearity diagnostics, including pairwise correlations and variance inflation factors (VIFs), were used to identify and avoid retaining highly correlated variables. The D/S/A framework further mitigates multicollinearity by iteratively removing redundant predictors during the deletion and substitution steps, favoring parsimonious models that maximize predictive performance while maintaining interpretability. No algorithms of temporal extensions to the predictors were applied during the modeling process. The modeling results, however, were applied to all the predictors across all the years to predict daily NO_2, $PM_{2.5}$ and O_3 concentrations for the 1989-2021 period.

Results

D/S/A integrated LUR models covering the available observational periods.

Table 3-5 present the daily LUR models, capturing the available observational time periods for NO_2, $PM_{2.5}$, and O_3. In the case of NO_2 (Table 3), the consistent year-after-year decline in concentrations observed during the study period was reflected in the variable "year", and this could be attributed to the regulatory efforts to reduce traffic NO_2 emissions. The recurrent pattern of lower concentrations during weekends compared to weekdays suggests potential reductions in human activities on roadways. Additionally, the positive correlation between higher OMI-NO_2 values and increased NO_2 concentrations underscores the significance of remote sensing observations in capturing spatial variability. Traffic density emerged as a significant factor, as areas with greater vehicular activity exhibited greater NO_2 emissions and higher concentrations. Moreover, weather conditions played a crucial role, with higher relative humidity, wind speed, and temperature contributing to lower NO_2 concentrations. Conversely, increased precipitation was linked to higher NO_2 levels, highlighting the interplay between meteorological conditions and NO_2 dynamics. Residential areas were found to have lower NO_2 concentrations, as well as in the developed open spaces. Low and high-intensity developments, on the other hand, were associated with greater NO_2 concentrations, indicating the positive association of urban development with NO_2 levels. The availability of green spaces, indicated by higher vegetation index, shrub cover, and wetlands, recognized as pollution sinks was associated with lower NO2 concentrations. Conversely, a higher proportion of impervious surfaces was correlated with increased NO2 levels. Additionally, locations farther from ports displayed lower NO_2 concentrations, indicating elevated NO_2 levels near ports. The NO_2 model had an adjusted R^2 of 0.84 in variance explained.

For $PM_{2.5}$ (Table 4), throughout the study period, its concentrations consistently decreased, mirroring the trend observed for NO_2. The study identified a positive correlation between higher aerosol optical depth (AOD) values and elevated $PM_{2.5}$ concentrations, suggesting that increased aerosol presence in the atmosphere is associated with higher particulate matter levels. Increased traffic density emerged as a contributing factor to higher $PM_{2.5}$ concentrations, emphasizing the impact of vehicular emissions on air quality. Weather factors such as higher relative humidity, wind speed, and temperature were associated with lower $PM_{2.5}$ concentrations. Developed open spaces were linked to reduced $PM_{2.5}$ concentrations, and so were areas characterized by a higher vegetation index, shrub cover, barren land, and water bodies, emphasizing the role of natural features in mitigating air pollution. Barren land refers to areas that have little to no vegetation cover and is often characterized by exposed soil or rock[122]. Industrial land use, however, was associated with higher $PM_{2.5}$ concentrations, pointing to the impact of industrial activities on particulate matter emissions. In contrast to NO_2, greater residential areas were linked to higher $PM_{2.5}$ concentrations, potentially attributed to background concentrations. In densely populated regions, the increased density of housing, traffic, and other activities can lead to elevated $PM_{2.5}$

background concentrations. Additionally, the urban heat island effect and limited air circulation in residential areas can hinder the dispersion of pollutants, allowing background $PM_{2.5}$ levels to rise. Additionally, locations farther from the coast were associated with higher $PM_{2.5}$ concentrations, indicating a spatial relationship between proximity to the coast and particulate matter levels. The final $PM_{2.5}$ model had a predictive performance of 0.65.

In contrast to the patterns observed for NO_2 and $PM_{2.5}$, O_3 concentrations exhibited predominantly opposing relationships (Table 5). The variable "year" did not show a significant association with O_3 concentrations, indicating the absence of an annual trend in O_3 levels. Weekends were characterized by higher O_3 concentrations than weekdays, revealing a distinct opposite temporal pattern. Higher OMI-O_3 values were linked to greater O_3 concentrations, emphasizing the positive association of remote sensing observations with measured ozone levels. Surprisingly, greater traffic was associated with lower O_3 concentrations, suggesting a nuanced photochemical process (i.e., scavenger effect, see details in discussion of Figure 4) between vehicular emissions and ozone dynamics. Weather factors such as higher relative humidity, wind speed, and atmospheric pressure correlated with elevated O_3 concentrations, underscoring the influence of meteorological conditions on ozone levels. Land use patterns also played a role, with government & institutional, commercial, and waterbody areas associated with higher O_3 concentrations, while barren land, crops, and wetlands were linked to lower O_3 concentrations. Developed low, medium, and high-intensity developments were associated with lower ozone concentrations, suggesting potentially lower concentrations in urban areas. Low-intensity development includes areas with sparse residential or commercial buildings, such as small towns or suburban neighborhoods. Medium-intensity development encompasses areas with more concentrated buildings and infrastructure, typically found in denser suburban or urban areas with moderate residential and commercial activities. High-intensity development represents the most densely built areas, including central business districts and urban centers with significant residential, commercial, and industrial structures[122]. Moreover, greater distances from highways were associated with higher O_3 concentrations, highlighting a similar scavenger effect between proximity to highways and ozone levels. The final O_3 model had a predictive performance of 0.92.

Table 3. Daily NO_2 model covering available observational periods.

Coefficient	Estimates	std. Error	Statistic	P-Value
Year	-0.166543	0.002916	-57.117810	**<0.001**
Season [Fall]	365.054890	5.843852	62.468199	**<0.001**
Season [Spring]	361.935364	5.842652	61.947107	**<0.001**
Season [Summer]	361.980633	5.841784	61.964058	**<0.001**
Season [Winter]	365.626222	5.844764	62.556198	**<0.001**
Week [Weekend]	-2.980948	0.024376	-122.29108	**<0.001**
NO_2 from OMI	8.50E-16	3.63E-18	234.413151	**<0.001**
Vehicle Kilometers Traveled (VKT) (350m)	0.000083	0.000001	117.218964	**<0.001**
Minimum Relative Humidity (%)	-0.135210	0.000722	-187.14711	**<0.001**
Wind Velocity at 10m (m/s)	-1.079110	0.007579	-142.38620	**<0.001**
Minimum Temperature (K)	-0.052602	0.002938	-17.906301	**<0.001**
Precipitation (mm, daily total)	0.034221	0.002549	13.425262	**<0.001**
Roadway Area (ha) (50m)	0.422647	0.009005	46.934432	**<0.001**
Residential (ha) (350m)	-0.007278	0.000145	-50.070525	**<0.001**
Waterbody (ha) (50m)	-1.886457	0.049965	-37.755221	**<0.001**
Developed Open Space (ha) (50m)	-0.142732	0.009172	-15.561964	**<0.001**
Developed Low Intensity (ha) (400m)	0.010865	0.000159	68.359609	**<0.001**
Developed High Intensity (ha) (5000m)	0.000104	0.000001	73.034457	**<0.001**
Shrubs (ha) (3250m)	-0.000073	0.000002	-34.732591	**<0.001**
Wetlands (ha) (550m)	-0.033821	0.000833	-40.579188	**<0.001**
NDVI	-0.000149	0.000011	-13.104524	**<0.001**
Percent Impervious (%) (50m)	0.037625	0.000605	62.143918	**<0.001**
Distance to Ports (m)	-0.000002	0.000000	-10.090608	**<0.001**
Distance to Highway (m)	-0.000091	0.000003	-29.413285	**<0.001**
Observations	321297			
R^2 / R^2 adjusted.	0.836 / 0.836			

Table 4. Daily $PM_{2.5}$ model covering available observational periods.

Coefficient	Estimates	std. Error	Statistic	P-Value
Year	-0.139709	0.003115	-44.847889	**<0.001**
Season [Fall]	360.125186	6.244769	57.668292	**<0.001**
Season [Spring]	356.974440	6.245309	57.158809	**<0.001**
Season [Summer]	358.493294	6.246059	57.395114	**<0.001**
Season [Winter]	360.534093	6.244547	57.735832	**<0.001**
AOD (albedo)	0.044977	0.000221	203.299221	**<0.001**
Vehicle Kilometers Traveled (VKT) (350m)	0.000012	0.000001	16.793841	**<0.001**
Wind Velocity (m/s)	-1.239394	0.006771	-183.031784	**<0.001**
Minimum Temperature (K)	-0.239641	0.002586	-92.662860	**<0.001**
Minimum Relative Humidity (%)	-0.059829	0.000649	-92.242887	**<0.001**
Roadway Area (ha) (5000m)	0.000024	0.000002	13.114503	**<0.001**
Industrial (ha) (1850m)	0.000513	0.000024	21.714939	**<0.001**
Residential (ha) (850m)	0.001185	0.000029	41.076124	**<0.001**
Unknown Land Use (ha) (450m)	-0.002008	0.000150	-13.387931	**<0.001**
Agricultural (ha) (50m)	-0.311401	0.014300	-21.776931	**<0.001**
NDVI	-0.000394	0.000010	-39.979943	**<0.001**
Barren Land (ha) (3000m)	-0.001291	0.000013	-99.546262	**<0.001**
Barren Land (ha) (50m)	-0.982108	0.074570	-13.170308	**<0.001**
Shrub Land (ha) (200m)	-0.029789	0.000822	-36.232176	**<0.001**
Developed Open Space (ha) (4950m)	-0.000037	0.000002	-16.515144	**<0.001**
Waterbody (ha) (1750m)	-0.000578	0.000020	-29.560264	**<0.001**
Distance to Highway (m)	-0.000029	0.000003	-8.723557	**<0.001**
Distance to Coast (m)	0.000017	0.000000	88.728793	**<0.001**
Elevation (m)	-0.002428	0.000053	-46.003552	**<0.001**
Observations	633277			
R^2 / R^2 adjusted	0.652 / 0.652			

Table 5. Daily O_3 model covering available observational periods.

Coefficient	Estimates	std. Error	Statistic	P-Value
Season [Fall]	-14.458252	1.093975	-13.216251	**<0.001**
Season [Spring]	-7.053920	1.104199	-6.388269	**<0.001**
Season [Summer]	-11.324983	1.112992	-10.175261	**<0.001**
Season [Winter]	-15.288623	1.087035	-14.064514	**<0.001**
Week [Weekend]	1.485925	0.029979	49.566212	**<0.001**
O_3 from OMI	0.046811	0.000557	84.056414	**<0.001**
Vehicle Kilometers Traveled (VKT) (350m)	-0.000055	0.000001	-50.399347	**<0.001**
Vapor Pressure (kPa)	6.030856	0.020439	295.059906	**<0.001**
Minimum Temperature (K)	0.098247	0.003827	25.669391	**<0.001**
Wind Velocity at 10m (m/s)	0.540921	0.008509	63.571245	**<0.001**
Government & Institutional (ha) (1800m)	0.000064	0.000007	9.086732	**<0.001**
Commercial (ha) (3200m)	0.000012	0.000009	1.249078	0.212
Waterbody (ha) (700m)	0.009638	0.000155	62.234975	**<0.001**
Developed Low Intensity (ha) (200m)	-0.060691	0.000715	-84.924620	**<0.001**
Developed Medium Intensity (ha) (150m)	-0.110609	0.000777	-142.421056	**<0.001**
Developed High Intensity (ha) (100m)	-0.249156	0.001589	-156.768640	**<0.001**
Barren Land (ha) (250m)	-0.058906	0.001805	-32.632019	**<0.001**
Crops (ha) (5000m)	-0.000085	0.000001	-158.369440	**<0.001**
Wetlands (ha) (1600m)	-0.003933	0.000053	-73.970854	**<0.001**
Distance to Highway (m)	0.000020	0.000002	12.767938	**<0.001**
Observations	513030			
R^2 / R^2 adjusted	0.923 / 0.923			

Daily air pollution surfaces

Figure 4 shows the aggregated annual concentration surfaces of NO_2 for four decennial years, including 1990, 2000, 2010, and 2020. The spatial patterns clearly show the decrease in NO_2 concentrations throughout the years, especially in the urban areas. To identify degrees of reduction throughout California, we used regulatory monitors for NO_2, $PM_{2.5}$, and O_3 (Figure 3) to identify average decennial concentrations for the State. This approach is reasonable given the state regulatory monitors are designed to ensure comprehensive spatial coverage, capturing the diverse environmental conditions across the state, including coastal, inland, and mountainous regions. By incorporating monitoring points from both urban and rural areas, it enables the examination of the urban-rural gradient in air pollution. These holistic statewide air quality monitors also allow for the identification of spatial patterns, hotspots, and potential disparities in pollution concentrations. Though some points are duplicated due to multiple pollutants being measured at the same time, they reflect the importance of those points in geographic placement strategies. Moreover, utilizing data from 1410 monitoring sites enhances the statistical robustness of the analysis, providing a more accurate assessment of statewide air pollution levels. Using those 1410 locations, we found that the average NO_2 concentrations decreased from 18.1 ppb in 1990 to 14.1 ppb in 2000, and decreased to 9.7 ppb in 2010 and 8.0 ppb in 2020. For $PM_{2.5}$, similar trends were identified for the four decennial years but with a much smaller decrease (Figure 5). A striking change in 2020 was that the $PM_{2.5}$ levels increased significantly in Central Valley while other places decreased, especially in Los Angeles, which experienced the greatest decline. We suspect the significant increase in $PM_{2.5}$ levels in Central Valley in 2020 was due to the significant impact of wildfires.[123] Using the locations of the 1410 regulatory monitors, we found that the average $PM_{2.5}$ concentrations decreased from 14.2 $\mu g\ m^{-3}$ in 1990 to 12.0 $\mu g\ m^{-3}$ in 2000, and further decreased to 9.9 $\mu g\ m^{-3}$ in 2010 but increased to 12.2 $\mu g\ m^{-3}$ in 2020. The increase in wildfire frequency and intensity in California[123–125] will further increase $PM_{2.5}$ levels, though regulatory actions have significantly reduced traffic and industry-related $PM_{2.5}$.

For O_3 (Figure 6), we did not see any apparent trend, but we did identify those urban metropolitan areas, such as the San Francisco Bay and Los Angeles Metro, had relatively lower O_3 concentrations compared to rural areas. This is very likely due to the O_3 scavenger effect[126]. The scavenger effect involves the removal or reduction of ozone from the atmosphere due to the presence of specific pollutants or conditions. These pollutants can act as scavengers by reacting with ozone molecules, leading to a decrease in overall ozone concentrations. Common scavengers of ozone include nitrogen oxides (NOx), carbon monoxide (CO), volatile organic compounds (VOCs), and particulate matter. In urban environments, where these pollutants are often abundant due to human activities such as combustion processes and industrial emissions, the scavenger effect can be more pronounced. Nitrogen oxides, particularly NO_2, can react with ozone in the presence of sunlight to form nitric oxide (NO) and oxygen (O_2). This process reduces the overall ozone levels in the atmosphere. VOCs and carbon monoxide can also participate in ozone-depleting reactions. These compounds can undergo photochemical reactions

that consume ozone while generating other pollutants. Using a total of 1410 spatial points from regulatory monitors, we found that the overall O_3 level did not change significantly through those four decennial years: the average O_3 concentrations decreased from 38.2 ppb in 1990 to 37.8 ppb in 2000, and slightly increased to 38.1 ppb in 2010 and 39.3 ppb in 2020.

Further, we provided daily air pollution surfaces for NO_2, $PM_{2.5}$, and O_3 for January 1st, 2019, and compared them with the corresponding nearest centennial annual surfaces (Figures 4-6). We found that for NO_2, the daily surface closely matched the spatial patterns of the annual surface. For $PM_{2.5}$, the patterns were also similar, though there was a significant increase in the Sierra region (eastern part of the map), suggesting a potential impact from wildfires. For O_3, while the general patterns were consistent in Northern California, the LA metropolitan area in Southern California showed higher O_3 concentrations on the daily map, which were less prominent in the annual data. Given the significant seasonal variability of O_3, we also compared seasonal means and observed notable differences both between seasons and relative to the annual averages. These comparisons indicate that while spatial patterns were largely consistent from daily to annual concentrations, there were notable differences in daily spatial patterns, particularly for $PM_{2.5}$ and O_3, likely due to the impact of temporal factors like wildfires and weather.

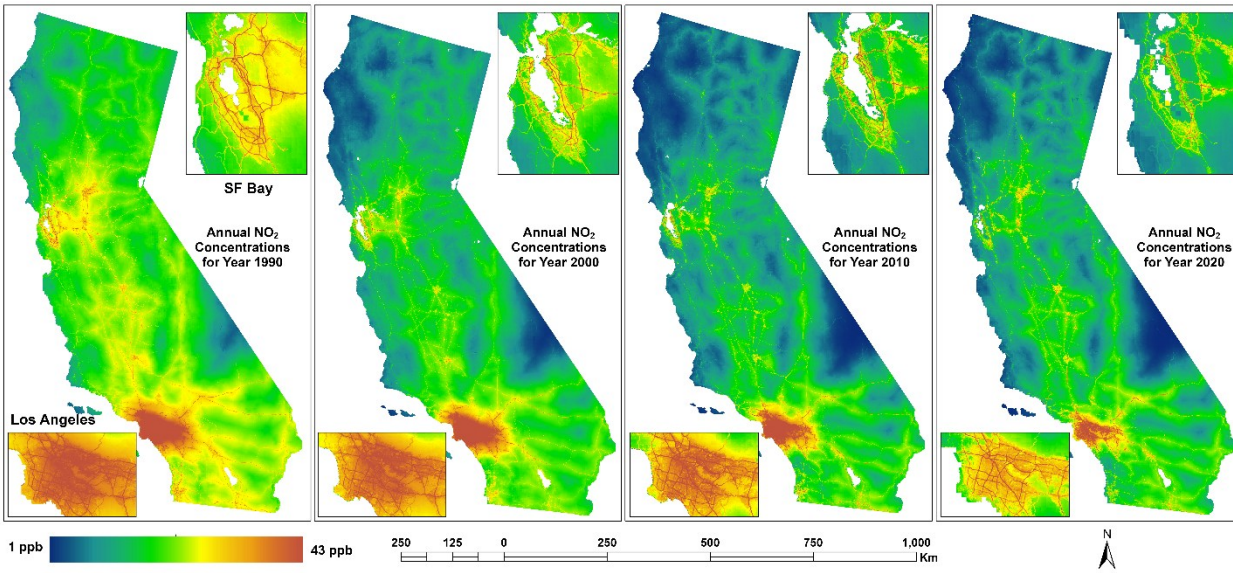


Figure 4. Decennial years of NO_2 surfaces among the over 30- years study period.

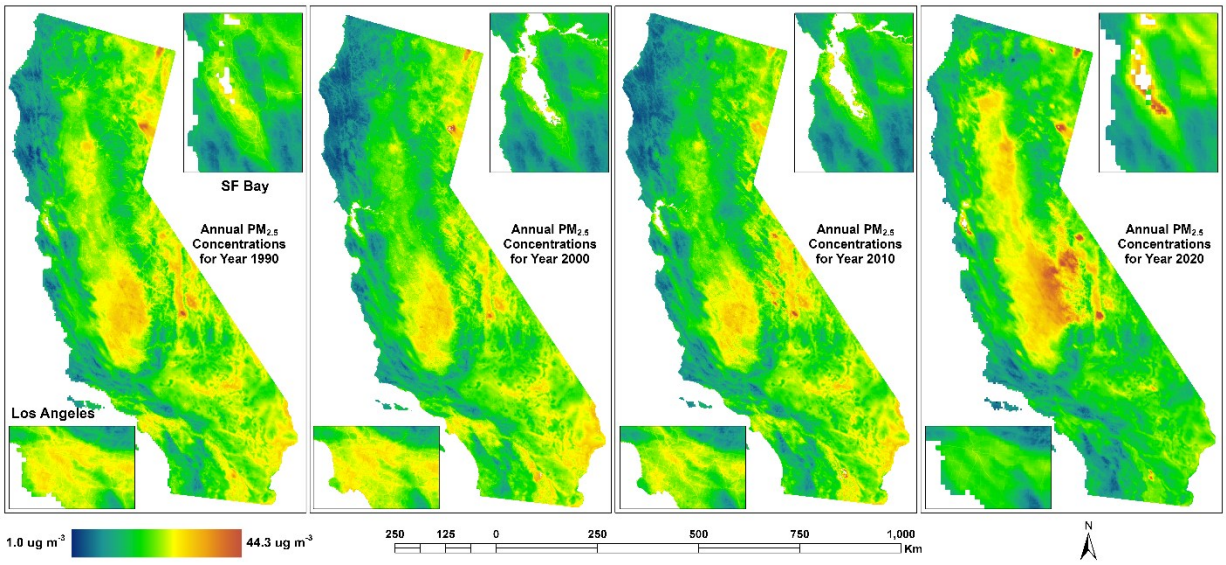


Figure 5. Decennial years of $PM_{2.5}$ surfaces among the over 30- years study period.

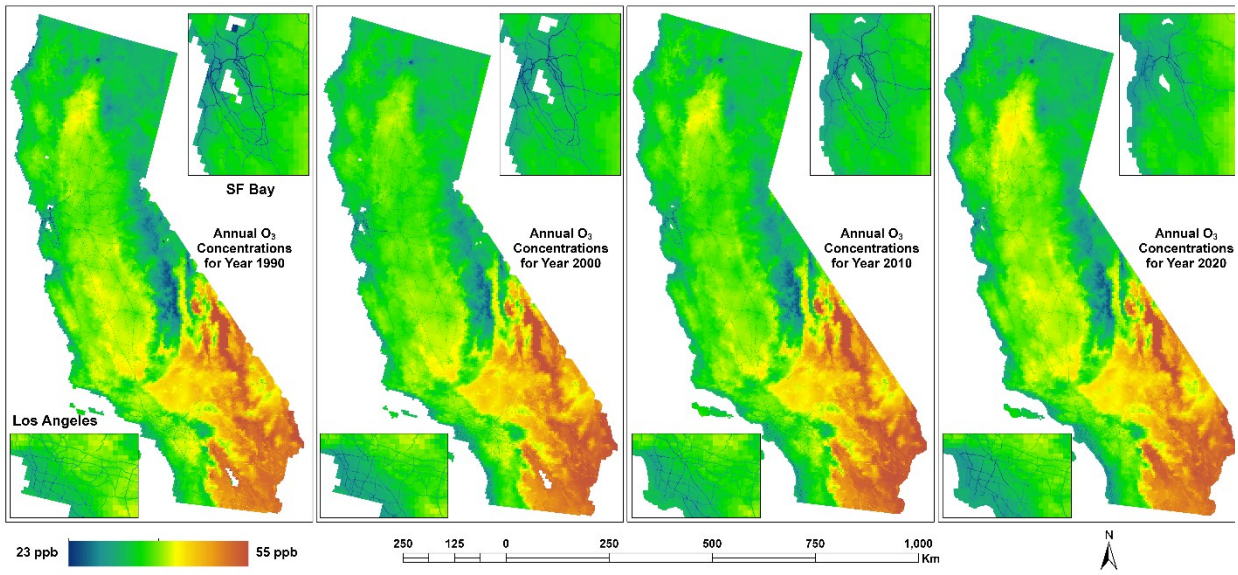


Figure 6. Decennial years of O_3 surfaces among the over 30- years study period.

Task 3. Develop air pollution models and surfaces for air toxics

Methodology

We applied the same deletion/substitution/addition machine learning LUR algorithms as used in the criteria pollutants for California. Air toxics, which include hazardous air pollutants (HAPs) such as chromium, nickel, lead and zinc, as well as volatile organic compounds (VOCs) like 1,3-butadiene and benzene, were incorporated in our air pollution modeling.

Description of air toxics regulatory monitoring data

The air toxics data used in our analysis was obtained from the CARB's speciation data and span from 1996 to 2021; however, measurements are sparse and irregular over time. This comprehensive dataset provides information on several key pollutants that are monitored for their hazardous effects on health and the environment. Benzene, a volatile organic compound commonly emitted by vehicle exhaust and industrial processes, was monitored at 54 unique sites. A total of 384 observations were recorded, providing insights into the spatial and temporal variations in benzene concentrations. These observations represent individual sampling events rather than continuous monitoring and are concentrated largely within the past decade, with some calendar years having no measurements at all. Importantly, this data density is consistent with, and in fact exceeds, what is commonly used in traditional annual LUR studies, which are often developed using a single measurement per site. Similarly, 1,3-butadiene, another VOC associated with vehicular emissions and industrial activities, was measured at 91 unique sites with 899 total observations, the highest number of sites among the pollutants in this dataset. Chromium, a heavy metal linked to industrial processes like metal plating and combustion, was monitored at 46 unique sites, with 366 total observations. Nickel, often associated with industrial activities such as metal smelting, was also measured at 46 unique sites, with the same number of observations as chromium. Lead, once commonly emitted from leaded gasoline and various industrial activities, was similarly monitored at 46 unique sites, with a total of 366 observations. Zinc, another heavy metal commonly released by industrial processes and vehicle emissions, was monitored at the same 46 sites, with 366 total observations. While often associated with industrial processes and vehicle emissions, Zinc may also have a significant contribution from agricultural activities and forest management practices. In agriculture, zinc can be released into the air as a result of the use of fertilizers and pesticides that contain zinc compounds. Additionally, soil erosion, which can occur in agricultural settings, may contribute to airborne zinc particles. In forest management, the burning of biomass or the use of zinc-containing pesticides for pest control can also contribute to zinc emissions.

Development of potential predictors for air toxics modeling

Pollutant specific emissions

For this project, we utilized emissions estimates from California's toxics emissions inventory, covering stationary sources from 1996 to 2021. These emissions data encompass both point

sources reported by facility operators or air districts under the Air Toxics “Hot Spots” Program (AB 2588) and aggregated point sources estimated by the California Air Resources Board (CARB) and local air districts. This comprehensive dataset provided a long-term perspective on emissions trends and variability. To ensure accurate spatial representation, point source locations were geocoded using ArcGIS geocoding services based on the street addresses provided. This step allowed for precise mapping of emission sources, enhancing the spatial fidelity of the dataset. In addition to the individual air toxics emissions, we included total organic gases (TOG) and reactive organic gases (ROG) data from CARB's emissions inventory. These broader organic compound categories complemented the detailed toxics data, providing a more holistic view of emissions relevant to air quality and exposure assessments.

All emission data were transformed into annual raster formats with a resolution of 100 m. This high-resolution rasterization process enabled a detailed spatial representation of emissions, capturing their distribution and intensity over time. The resulting rasters served as key inputs for developing annual air toxics models and pollutant concentration surfaces. By integrating this detailed emissions data into the modeling process, we were able to use spatially and temporally resolved predictors to better understand and assess air toxics distributions.

Remote sensing Sentinel bands and ratios

In our LUR models for air toxics, Sentinel-2 spectral bands and their ratios are used as predictors to account for the spatial distribution of surface characteristics that influence pollutant levels[127]. These bands, ranging from visible to shortwave infrared (SWIR) wavelengths, capture variations in land cover, vegetation, and urbanization, which correlate with air pollutant concentrations.

For lead, significant coefficients were observed by[127] in the Blue (B2) band and the band ratio B6/B8, suggesting that these features effectively capture spatial variability associated with sources or sinks of lead pollution. Specifically, the Blue band (B2) positively correlates with lead concentrations, indicating a potential link with urban or industrial surface characteristics. The ratio B6/B8 is also significant, highlighting differences between vegetation and urban structures that may influence the spatial distribution of lead.

For zinc, key predictors included the Green (B3) band and band ratios B3/B8 and B6/B8. The Green band positively correlates with zinc, potentially reflecting surface characteristics associated with urban vegetation or metallic surfaces that influence zinc deposition. The negative coefficient of B3/B8 suggests an inverse relationship with vegetation density, while B6/B8 indicates that NIR-based spectral differences capture spatial variation in zinc pollution.

Overall, the integration of Sentinel-2 spectral data enhances the spatial resolution of air pollution modeling by leveraging the detailed information on land cover and surface characteristics to improve the prediction of air toxic distributions.

Traditional LUR predictors

For the development of our LUR models for nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$), and ozone (O_3), we utilized a wide range of integrated, comprehensive data sources that provided crucial spatial and temporal information. These sources included traffic data, land use and land cover data, meteorological conditions, vegetation dynamics, and satellite data. Please refer to the criteria pollutants for the predictors considered. Now, we are extending the use of these same predictors to model air toxics, ensuring a robust framework for understanding the spatial distribution of hazardous pollutants.

Traffic data are a critical input in air pollution modeling, particularly for pollutants like NO_2 and $PM_{2.5}$, which are heavily influenced by vehicle emissions. This data includes information on traffic volume and road networks. Traffic data helps identify areas with high vehicular emissions, which are significant contributors to local air quality, especially in urban environments. By incorporating this data into the models, we can account for the impact of transportation-related emissions on air toxics, as pollutants like benzene, 1,3-butadiene, and other air toxics are often linked to motor vehicle exhaust.

Land use and land cover data provide essential information about how different types of land cover and land use activities influence air quality. Land use refers to how land is utilized (e.g., residential, industrial, commercial, or agricultural), while land cover pertains to the physical surface of the land (e.g., urban areas, forests, grasslands, or water bodies). This data allows us to understand the relationship between human activity, land transformation, and pollution patterns. In the case of air toxics, land use and land cover data can help pinpoint areas of high pollution exposure, such as industrial zones, high-density residential areas, or regions affected by agricultural practices.

Meteorological conditions are a crucial factor in the dispersion and transformation of air pollutants. We incorporated data on temperature, wind speed, wind direction, humidity, and atmospheric pressure, which all influence how pollutants travel, dilute, and react in the atmosphere. For example, wind patterns can carry pollutants from high-emission areas to other regions, while temperature and humidity influence the formation of secondary pollutants such as ozone. In air toxics modeling, meteorological data helps account for how local weather conditions can impact the spread and concentration of hazardous pollutants, providing a more dynamic and accurate representation of air quality.

Vegetation dynamics also play a significant role in air pollution modeling. Vegetation can act as both a sink and a source for certain pollutants. For instance, plants can absorb some pollutants through their stomata, reducing local pollutant concentrations. Conversely, land management practices such as deforestation, urbanization, or changes in vegetation cover can alter the natural processes that help mitigate air pollution. By including vegetation dynamics in our models, we can account for how vegetation types and coverage influence the dispersion of air toxics, particularly in rural or suburban areas where vegetation plays a more prominent role in air quality.

Finally, satellite data provides a powerful tool for capturing large-scale spatial patterns of land cover, vegetation, and even pollutant concentrations. Satellite imagery, often available at high resolution, can be used to monitor changes in land use, track vegetation dynamics, and even estimate pollutant levels from space. For example, vegetation indices derived from satellite data, such as the Normalized Difference Vegetation Index (NDVI), can offer insights into the density and health of vegetation in a given area. Furthermore, satellite-based remote sensing can assist in monitoring emissions from point sources, providing another layer of data to refine the modeling of air toxics.

Deletion/Substitution/Addition (D/S/A) LUR modeling techniques

For the annual air toxics modeling, we applied a similar approach to the one used in developing daily air pollution models for NO_2, $PM_{2.5}$, and O_3. However, as air toxics data is only available from CARB's speciation data and EMFAC emissions data are annual, our focus shifted to developing annual surfaces rather than daily models. This shift in temporal resolution necessitated some adjustments to the modeling process, but the core methodology remained consistent.

The modeling for air toxics follows the D/S/A algorithm, which begins with a base model, typically an intercept-only model, unless otherwise specified. The algorithm then iteratively modifies the model by adding, deleting, or substituting terms to improve its predictive performance. In each iteration, potential model modifications—such as incorporating polynomial terms or testing different combinations of predictors—are evaluated against a predefined criterion, typically focused on improving cross-validated error or another performance metric. This iterative process tests various combinations of terms, retaining only those that result in the greatest improvement in model accuracy. The algorithm continues to iterate until no significant improvement is observed in model performance, based on predefined stopping criteria. These criteria could include reaching a threshold for improvement in cross-validated R-squared or completing a maximum number of iterations, which was set to 15 in our previous work. Once the algorithm converges, the optimal model is selected, representing the best balance between model complexity and predictive accuracy. To maintain model interpretability, we restricted the predictors to linear terms, avoiding interaction effects.

For cross-validation, the same approach used in the daily LUR models for NO_2, $PM_{2.5}$, and O_3 was applied, where data was randomized and divided into 10 equal folds. This ensures a balanced mix of data in each fold, helping to avoid overfitting while maintaining model robustness. This cross-validation process ensured the air toxics models were optimized for predictive accuracy across different regions and temporal scales.

In summary, while the transition to annual air toxics modeling required some adjustments to the temporal framework, the D/S/A algorithm and the methodology for selecting optimal predictors remained largely the same as for NO_2, $PM_{2.5}$, and O_3. This consistent approach allowed us to develop reliable models for predicting the distribution of hazardous air pollutants across time and space.

Results

The model for benzene (Table 6) had an Adjusted R^2 of 0.806, indicating that the model accounts for 80.6% of the variability in the data. This relatively high explanatory power demonstrates that the model captures the dominant spatial determinants of benzene levels and provides a statistically robust representation of benzene exposure patterns across California.. The year variable, with a negative coefficient, suggests that benzene concentrations have generally declined over the years. Among the land use predictors, developed open space within 1500 m has a negative and significant association with benzene levels. Developed high-intensity land use at a 5000 m buffer distance shows a positive effect on benzene. Further, the developed medium-intensity areas within 50 m show a significant positive effect, indicating that more densely developed areas tend to have higher benzene levels. Certain wetland areas also show significant relationships with benzene concentrations. For example, wetlands within 1500 m are negatively associated with benzene, suggesting that wetlands may help mitigate benzene levels, possibly due to vegetation and land cover features. However, wetlands within 4200 m show a positive relationship with benzene, reflecting a more distant influence of wetland areas on benzene concentrations. Urban-related variables, such as impervious surface percentage within 1250 m and residential areas within 2550 m, have a positive association with benzene levels, indicating that increased urbanization and impervious surfaces contribute to higher benzene concentrations. On the industrial front, industrial areas within 150 m are negatively associated with benzene, which could be due to specific regulatory controls or other factors like plume effect. However, industrial areas at larger distances (4200 m) have a positive effect. Key environmental variables, such as wind velocity at 10 m, have a strong negative relationship with benzene concentrations, suggesting that higher wind speeds may help disperse benzene and lower local concentrations. Maximum temperature shows a positive relationship with benzene, indicating that warmer temperatures might lead to higher benzene levels, which is consistent with the fact that benzene is a volatile organic compound that can increase in warmer conditions.

The model for 1,3-butadiene (Table 7) had an adjusted R^2 of 0.619, demonstrating a moderate ability to explain the variability in 1,3-butadiene levels. The dataset includes 899 observations from various sites, ensuring a broad representation of environmental conditions. A clear decreasing trend in 1,3-butadiene concentrations over time highlights the impact of regulations and improved emission controls. Industrial and vehicular activities are major contributors to ambient levels. Proximity to industrial areas is associated with higher concentrations, confirming the role of industrial emissions as a primary source. Interestingly, industrial activity at slightly larger distances shows a negative association, potentially due to dispersion effects. Vehicle-related predictors, such as vehicle kilometers traveled, also play a significant role, reaffirming the importance of transportation emissions. Land use and vegetation have notable effects. Grasslands and areas with higher vegetation indices are linked to lower concentrations of 1,3-butadiene, suggesting that vegetation helps mitigate pollution levels, possibly through pollutant deposition or reduced emissions. In contrast, open land near measurement sites is associated with higher concentrations, which could reflect emissions from unregulated sources. Organic gases are also key predictors. ROG within close proximity are linked to lower 1,3-butadiene concentrations, likely due to chemical reactions or differing sources. However, at larger distances, ROG levels are positively associated with 1,3-butadiene concentrations, indicating more complex spatial and transport processes. TOG shows a positive relationship, consistent with their role in combustion-related emissions. Meteorological factors, including wind speed and temperature, significantly influence 1,3-butadiene concentrations. Higher wind speeds are associated with lower concentrations, as dispersion reduces pollutant buildup. Conversely, higher temperatures are linked to increased concentrations, potentially due to enhanced emissions or chemical reactions.

The chromium (Table 8) model had an adjusted R^2 of 0.758, indicating strong explanatory power. Based on 366 observations, the analysis highlights several key predictors related to land use, industrial activity, and other environmental characteristics. Land-use variables show varying relationships with chromium concentrations. Developed open spaces close to measurement sites are associated with lower chromium levels, suggesting their role in reducing pollution exposure. However, at greater distances, these spaces show a positive association, indicating that chromium emissions may disperse from distant developed areas. Similarly, low-intensity developed areas near the sites are associated with higher chromium levels, reflecting the potential impact of moderate urban activities. At larger distances, these areas have a mitigating effect, which may reflect dispersion effects or reduced direct emissions. Natural land uses also play a role. Forested and shrub lands show positive associations with chromium concentrations, which could indicate deposition of airborne chromium from other sources rather than local emissions. These findings emphasize the complex role of vegetation in pollutant dynamics. Built environmental characteristics, such as impervious surfaces and residential areas, are significant contributors to chromium levels. Higher percentages of impervious surfaces in proximity to measurement sites are strongly linked to elevated chromium concentrations, consistent with the role of urbanized

areas in generating and retaining pollution. Residential areas also contribute positively, suggesting the influence of household and localized activities. Industrial activities are significant predictors of chromium levels. Proximity to industrial areas is associated with higher chromium concentrations, affirming the impact of industrial emissions. Interestingly, industrial activity at slightly greater distances has a negative association, possibly reflecting dispersion effects.

The nickel (Table 9) model explains a substantial portion of the variability in ambient nickel concentrations, with an adjusted R^2 of 0.698 based on 366 observations. The results indicate several significant predictors related to temporal trends, land use, atmospheric conditions, and environmental factors. Year shows a negative association with nickel levels, reflecting a decreasing trend over time, possibly due to improved regulations or reduced industrial emissions. Aerosol Optical Depth (AOD), a measure of atmospheric pollution, is positively associated with nickel concentrations, suggesting that higher aerosol levels contribute to increased ambient nickel. The Sentinel reflectance ratio, related to surface characteristics, has a negative effect, indicating that lower reflectance is linked to reduced nickel concentrations, potentially reflecting differences in land cover or emissions sources. Land use features show varying impacts depending on the type and proximity of the area. Developed open spaces demonstrate mixed effects; closer distances are associated with higher nickel levels, while intermediate distances exhibit a negative association, likely reflecting dispersion patterns or land-use intensity. High-intensity developed areas, particularly those very close to measurement locations, exhibit a strong positive association, likely reflecting emissions from industrial or urban activities. Similarly, cultivated land at greater distances shows a positive relationship, indicating contributions from agricultural regions, possibly due to fertilizer use or soil disturbance. Vegetative features, such as tree canopy cover near the measurement sites, show a negative association with nickel concentrations, suggesting that vegetation can play a role in mitigating pollution levels. The location category variable (with values 1 to 3), representing proximity to roadways, has a significant negative association with nickel concentrations. Locations within 500 m of highways (category 1) and those within 300 m of major roadways (category 2) generally experience higher nickel levels compared to control areas far from highways and major roadways (category 3). This pattern reflects the contribution of vehicular emissions to nickel concentrations.

The lead model (Table 10), with an adjusted R^2 of 0.585 based on 366 observations, captures key factors influencing ambient lead concentrations. Year exhibits a significant negative association with lead levels, indicating a decline over time, likely reflecting the effectiveness of regulatory measures such as the phase-out of leaded gasoline and stricter industrial emissions controls. Sentinel band 11 reflectance at 100 m, representing specific surface properties, shows a positive association with lead concentrations, suggesting that reflectance characteristics of certain surfaces may influence the deposition or re-emission of lead. Land use and land cover variables reveal complex spatial relationships with lead concentrations. Developed open space at intermediate distances (1750 m) has a significant negative association, while developed low-

intensity areas within 400 m exhibit a strong positive effect, highlighting the influence of urbanization and human activities on lead levels. High-intensity developed areas at farther distances (1850 m) also show a positive association, suggesting contributions from densely built environments. Industrial areas within 50 m have a substantial positive impact, underlining their role as key sources of lead pollution, likely from manufacturing processes or emissions. Cultivated land demonstrates mixed effects based on distance, with negative associations at closer proximities (1300 m) and positive effects at farther distances (1700 m and 4100 m). These results may reflect the combined influence of agricultural practices, including the historical use of lead-based pesticides, and spatial patterns of atmospheric lead deposition. Tree canopy cover within 150 m exhibits a significant negative association, indicating that vegetation may help reduce lead concentrations by capturing airborne particles. Wind velocity at 10 m is negatively associated with lead levels, likely due to enhanced dispersion of pollutants under windy conditions.

The zinc model (Table 11), with an adjusted R^2 of 0.902, demonstrates a highly robust explanation of spatial and temporal variation in ambient zinc concentrations. Based on 366 observations, the model highlights several predictors related to land use and land cover, water features, and meteorological influences, as well as the influence of time and anthropogenic activities. The year variable shows a significant positive association with zinc levels, indicating an increasing trend in concentrations over time. This could reflect growing zinc emissions from activities such as industrial production, urban development, agricultural use and forest management. Land use variables reveal the substantial impact of urbanization and agricultural practices on zinc concentrations. Developed open space has scale-dependent effects, with a significant positive association at 550 m and 4900 m, but a negative effect at 1650 m. This pattern may reflect the spatial heterogeneity of zinc sources, including construction activities, vehicular emissions, and material weathering in urban environments. Developed low-intensity areas within 100 m show a particularly strong positive association with zinc levels, likely reflecting contributions from residential land use, such as roofing materials, paints, and vehicle-related emissions. Industrial land within 300 m also has a significant positive effect, further underscoring the role of industrial activities as major contributors to zinc pollution. Agricultural and forested areas are also significant predictors, pointing to the dual influence of agricultural use and forest management on zinc concentrations. Cultivated land within 1300 m has a significant negative association, potentially due to localized zinc absorption by crops or soil processes. However, cultivated land at farther distances, such as 2750 m, has a pronounced positive association, reflecting agricultural runoff and emissions contributing to regional zinc deposition. Forest land within 100 m shows a positive association, possibly linked to forest management practices, including the use of zinc-containing fertilizers or burning of biomass (e.g., wildfires). Similarly, shrub land at 1750 m and tree canopy at 3400 m positively influence zinc levels, suggesting that vegetated areas may play a role in trapping or re-emitting zinc through biogeochemical processes. Other significant predictors include roadway areas within 1050 m, which are positively associated with zinc concentrations, likely due to tire and brake wear, as well as roadway dust resuspension. Commercial land use shows mixed effects, with a

negative association at 1550 m and a positive association at 2850 m, reflecting variations in commercial activities and their contribution to zinc emissions.

Table 6. The Benzene (ppb – parts per billion) annual land use regression model for the State of California.

Predictors	**Estimates**	**Statistic**	**p**
(Intercept)	71.1444530829	6.460868056	**<0.001**
Year	-0.0460644573	-8.187001384	**<0.001**
Developed open space (ha) (1500m)	-0.0003038868	-9.20768114	**<0.001**
Developed low-intensity (ha) (50m)	0.1046491564	6.108446398	**<0.001**
Developed low-intensity (ha) (150m)	0.0095996898	2.581800578	**0.01**
Developed low-intensity (ha) (400m)	-0.0040680773	-7.39114508	**<0.001**
Developed medium-intensity (ha) (50m)	0.0750590130	9.494233321	**<0.001**
Developed high-intensity (ha) (100m)	0.0091059982	3.24671851	**0.001**
Developed high-intensity (ha) (5000m)	0.0000233851	10.09621313	**<0.001**
Barren land (ha) (3650m)	0.0001646350	1.513935759	0.131
Wetlands (ha) (1500m)	-0.0006293209	-4.169578405	**<0.001**
Wetlands (ha) (4200m)	0.0000765548	5.221485589	**<0.001**
Percent impervious (%) (1250m)	0.0049819000	6.492024697	**<0.001**
Residential (ha) (2550m)	0.0001292767	15.79280818	**<0.001**
Commercial (ha) (400m)	0.0007969469	4.086255787	**<0.001**
Open land (ha) (4050m)	0.0000447860	12.98661681	**<0.001**
Agricultural (ha) (650m)	0.0003945438	2.859741801	**0.004**
Industrial (ha) (150m)	-0.0062452995	-2.986573997	**0.003**
Industrial (ha) (4200m)	0.0000341314	8.618888805	**<0.001**
Roadway area (ha) (400m)	0.0027072548	7.338750856	**<0.001**
Wind velocity at 10m (m/s)	-0.4824059638	-8.830365082	**<0.001**
Maximum temperature (K)	0.0740398482	8.81029949	**<0.001**
Observations	384		
R^2 / R^2 adjusted	0.817 / 0.806		

Table 7. The 1,3 Butadiene (ppt - parts per trillion) annual land use regression model for the State of California.

Predictors	**Estimates**	**Statistic**	**p**
(Intercept)	31117.6648826934	11.64674359	**<0.001**
Year	-21.1249730190	-19.38089739	**<0.001**
Barren land (ha) (4600m)	-0.0469201251	-2.167769104	**0.03**
Grass land (ha) (500m)	-0.6775093539	-5.129792766	**<0.001**
Normalized difference vegetation index (NDVI)	-0.0424093861	-4.420047065	**<0.001**
Reactive organic gases (ROG in Kg) (350m)	-25.1894794709	-4.523940379	**<0.001**
Reactive organic gases (ROG in Kg) (800m)	-2.9373635888	-2.93857921	**0.003**
Reactive organic gases (ROG in Kg) (1750m)	1.5820924780	9.938395905	**<0.001**
Reactive organic gases (ROG in Kg) (5000m)	0.0789423273	3.453985634	**0.001**
Total organic gases (TOG in Kg) (450m)	19.5173083814	6.458797477	**<0.001**
Industrial (ha) (100m)	25.0402701651	6.830834645	**<0.001**
Industrial (ha) (250m)	-4.7079719446	-7.984294052	**<0.001**
Open land (ha) (50m)	16.0480573790	2.957217315	**0.003**
Open land (ha) (3100m)	0.0049091507	2.688184263	**0.007**
Unknown land use (ha) (1150m)	-0.1412559852	-4.712224558	**<0.001**
Wind velocity at 10m (m/s)	-104.8729264054	-5.294000198	**<0.001**
Minimum temperature (K)	22.4987766572	3.637364344	**<0.001**
Maximum temperature (K)	18.9826049436	4.543141132	**<0.001**
Distance to ports (m)	0.0007354483	4.515508873	**<0.001**
Vehicle kilometer traveled (VKT) (350m)	0.0012401077	3.276795231	**0.001**
Observations	899		
R^2 / R^2 adjusted	0.627 / 0.619		

Table 8. The Chromium (μg m^{-3} – microgram per cubic meter) annual land use regression model for the State of California.

Predictors	**Estimates**	**Statistic**	**p**
(Intercept)	-0.0009652030	-3.421993114	**0.001**
Developed open space (ha) (750m)	-0.0000026233	-5.148937014	**<0.001**
Developed open space (ha) (4950m)	0.0000001187	7.855805639	**<0.001**
Developed low-intensity (ha) (750m)	0.0000011149	4.834187597	**<0.001**
Developed low-intensity (ha) (4350m)	-0.0000000776	-5.356293607	**<0.001**
Forest land (ha) (1700m)	0.0000000804	3.278770817	**0.001**
Shrub land (ha) (3500m)	0.0000000227	3.449440594	**0.001**
Percent impervious (%) (700m)	0.0000542740	13.2018244	**<0.001**
Residential (ha) (250m)	0.0000033797	2.733756416	**0.007**
Industrial (ha) (50m)	0.0150493090	6.183608152	**<0.001**
Industrial (ha) (100m)	-0.0006743840	-6.219904697	**<0.001**
Industrial (ha) (1850m)	-0.0000007260	-5.478485644	**<0.001**
Observations	366		
R^2 / R^2 adjusted	0.765 / 0.758		

Table 9. The Nickel (μg m^{-3} – microgram per cubic meter) annual land use regression model for the State of California.

Predictors	**Estimates**	**Statistic**	**p**
(Intercept)	0.0925145525	4.818773481	**<0.001**
Year	-0.0000457830	-4.800933522	**<0.001**
Aerosol Optical Depth (AOD)	0.0000042343	3.156966846	**0.002**
Sentinel reflectance ratio of band 3 to band 8 (2200m)	-0.0005806226	-2.149175643	**0.032**
Water (ha) (600m)	0.0000378123	2.45276961	**0.015**
Developed open space (ha) (750m)	0.0000012524	2.699465104	**0.007**
Developed open space (ha) (1650m)	-0.0000013001	-8.293587973	**<0.001**
Developed open space (ha) (4800m)	0.0000001079	8.329239371	**<0.001**
Developed low-intensity (ha) (150m)	-0.0000153411	-3.105717724	**0.002**
Developed low-intensity (ha) (550m)	0.0000046813	7.733307247	**<0.001**
Developed low-intensity (ha) (3000m)	-0.0000000709	-3.663505516	**<0.001**
Developed high-intensity (ha) (50m)	0.0000640834	3.593088787	**<0.001**
Developed high-intensity (ha) (800m)	0.0000012760	6.070259799	**<0.001**
Developed high-intensity (ha) (2100m)	-0.0000001901	-4.30779798	**<0.001**
Forest land (ha) (100m)	0.0000160534	1.938139917	0.053
Forest land (ha) (500m)	-0.0000003772	-1.515190092	0.131
Cultivated land (ha) (5000m)	0.0000000275	6.504952433	**<0.001**
Tree canopy (%) (50m)	-0.0000134758	-3.157904608	**0.002**
Unknown land use (ha) (4500m)	0.0000000053	2.458329336	**0.014**
Daily precipitation (mm)	0.0000597377	1.649510182	0.1
Location category	-0.0002111528	-4.722682297	**<0.001**
Observations	366		
R^2 / R^2 adjusted	0.714 / 0.698		

Table 10. The Lead (µg m^{-3} – microgram per cubic meter) annual land use regression model for the State of California.

Predictors	**Estimates**	**Statistic**	**p**
(Intercept)	0.1786862844	2.853946155	**0.005**
Year	-0.0000879009	-2.826373007	**0.005**
Sentinel band 11 reflectance (%) (100m)	0.0000010686	3.048932125	**0.002**
Water (ha) (550m)	0.0000824566	1.507671915	0.133
Developed open space (ha) (1750m)	-0.0000010025	-4.125566514	**<0.001**
Developed low-intensity (ha) (400m)	0.0000067087	4.44248285	**<0.001**
Developed high-intensity (ha) (1850m)	0.0000002179	2.168973687	**0.031**
Shrub land (ha) (100m)	-0.0000184916	-1.493241436	0.136
Cultivated land (ha) (1300m)	-0.0000112322	-5.260807218	**<0.001**
Cultivated land (ha) (1700m)	0.0000063385	6.671948391	**<0.001**
Cultivated land (ha) (4100m)	0.0000002204	3.950092096	**<0.001**
Tree canopy (%) (150m)	-0.0000240071	-2.077729624	**0.038**
Residential (ha) (50m)	0.0001268974	1.42360566	0.155
Commercial (ha) (50m)	-0.0001516211	-1.779575967	0.076
Commercial (ha) (850m)	-0.0000035623	-2.816310915	**0.005**
Commercial (ha) (1600m)	0.0000011326	1.700822505	0.09
Industrial (ha) (50m)	0.0080694047	2.300571321	**0.022**
Daily precipitation (mm)	0.0004079364	3.204730598	**0.001**
Vapor pressure deficit (kPa)	-0.0006627179	-1.673587572	0.095
Wind velocity at 10m (m/s)	-0.0004811740	-2.813558467	**0.005**
Distance to ports (m)	-0.0000000019	-1.560162738	0.12
Observations	366		
R^2 / R^2 adjusted	0.608 / 0.585		

Table 11. The Zinc (μg m^{-3} – microgram per cubic meter) annual land use regression model for the State of California.

Predictors	Estimates	Statistic	p
(Intercept)	-0.7633161249	-4.717188522	**<0.001**
Year	0.0003761687	4.678444883	**<0.001**
Water (ha) (600m)	0.0026681203	7.642488406	**<0.001**
Water (ha) (650m)	-0.0024089350	-7.281817116	**<0.001**
Water (ha) (950m)	0.0003425548	10.06154186	**<0.001**
Water (ha) (1100m)	-0.0001080435	-9.297688355	**<0.001**
Developed open space (ha) (550m)	0.0000395784	7.932539305	**<0.001**
Developed open space (ha) (1650m)	-0.0000155901	-14.24917187	**<0.001**
Developed open space (ha) (4900m)	0.0000008815	9.760316747	**<0.001**
Developed low-intensity (ha) (100m)	0.0005624142	9.413001078	**<0.001**
Developed low-intensity (ha) (1100m)	0.0000029328	3.200835573	**0.001**
Forest land (ha) (100m)	0.0001372166	2.944495893	**0.003**
Shrub land (ha) (1750m)	0.0000006002	3.713286452	**<0.001**
Cultivated land (ha) (1300m)	-0.0000566748	-13.28121271	**<0.001**
Cultivated land (ha) (2750m)	0.0000139214	28.73397069	**<0.001**
Tree canopy (%) (3400m)	0.0001456838	5.016416366	**<0.001**
Commercial (ha) (1550m)	-0.0000126516	-11.18764605	**<0.001**
Commercial (ha) (2850m)	0.0000034115	5.848703078	**<0.001**
Industrial (ha) (300m)	0.0001951988	11.73936751	**<0.001**
Roadway area (ha) (1050m)	0.0000086079	7.391120595	**<0.001**
Observations	366		
R^2 / R^2 adjusted	0.907 / 0.902		

Task 4. Data acquisition of human subjects' data for 2010-2019

The UCB research team acquired three primary datasets to assess the impacts of air pollution on metabolic health outcomes across California from 2010 to 2019. These datasets include: (1) the CHIS data for diabetes incidence and medication use, (2) the HCAI data for diabetes-related emergency department (ED) visits and hospitalizations, and (3) the CDPH Vital Records for diabetes-related mortality. All datasets were obtained at the individual level, with CHIS and CDPH data including residential addresses and HCAI data linked at the five-digit ZIP code level. The CHIS data were acquired for years 2011-2019 for every two years and stored in the University of California, Los Angeles (UCLA) Center for Health Policy Research (CHPR). University of California, Berkeley (UCB) designed R code for CHPR staff to run the analysis. For the data provided by HCAI and CDPH, Institutional Review Board (IRB) approval was obtained to ensure the secure and ethical use of human subject data by authorized UCB researchers. We submitted applications to both the UCB Institutional Review Board (for reliance on State Committee for Protection of Human Subjects - CPHS) and the California Health and Human Services Committee, both of which reviewed and approved our research protocol. Following approval, we worked with HCAI and CDPH to acquire related data under a strict data-use agreement to protect confidentiality and the acquired data were stored on secure UCB Secure Research Data Center (SRDC) servers in compliance with the Health Insurance Portability and Accountability Act (HIPAA) and state requirements.

The CHIS dataset, the largest state health survey in the U.S., provided detailed, population-representative data on California residents. UCB identified survey participants diagnosed with T2D between 2011 and 2019. Variables extracted included latitude/longitude of home address, age, gender, race-ethnicity, insurance status, body mass index (BMI), smoking status, diabetes medication use (e.g., insulin injections or oral medications), and hemoglobin A1C checks. These data enabled the team to examine population-level diabetes incidence, treatment behaviors, and disparities across demographic groups and geographic areas.

The HCAI dataset provided comprehensive statewide records of T2D mellitus (ICD-9 code 250; ICD-10 code E11) for both ED visits and hospitalizations. Collected variables included patient five-digit ZIP code, date of admission and discharge, length of stay, age, gender, race-ethnicity, Elixhauser comorbidity index (derived from diagnosis codes), facility number, payer category, preferred language, principal procedure, and care type. Similarly, the CDPH Vital Records dataset captured mortality events where diabetes was listed as the primary or contributory cause of death. Variables included residential address, date of death, age, gender, race-ethnicity, smoking status, BMI, and insurance information. To further characterize diabetes-related deaths, UCB also obtained data on underlying causes of death when diabetes was listed as a secondary cause.

Task 5. Identify concentration-response relationships between air pollution exposures and five health endpoints

Task 5 was designed to comprehensively evaluate how air pollution exposures are associated with multiple diabetes-related health outcomes across different temporal scales and population subgroups. Specifically, this task aimed to (1) quantify exposure-response relationships using appropriate short-term and long-term exposure windows, (2) assess lagged exposure patterns to characterize current versus delayed effects, and (3) identify population groups exhibiting greater vulnerability. Long-term exposure analyses were conducted for diabetes incidence, medication use (using CHIS data), and diabetes-related mortality (using CDPH Vital Records), while short-term exposure analyses were conducted for diabetes-related emergency department visits, hospitalizations, and length of stay using HCAI data.

Assigning air pollution exposure to locations of subjects

For CHIS, surveys were conducted biennially (approximately every two years). The daily air pollution exposure (NO_2, $PM_{2.5}$, O_3) was therefore aggregated to annual metrics and assigned retrospectively to each census tract back to 1989. The annual air toxics surfaces were assigned to the census tract in a way like the aggregated annual criteria pollutants. High-resolution (100 m) pollution surfaces were aggregated to census tracts using block-group population-weighted means to approximate population exposure within each tract. Exposures were assigned back to 1989 specifically because the development of T2D may have occurred long before survey years. One year before the year of incidence of a survey participant was treated as air pollution exposure for that participant. Incidence records with onset prior to 1990 were excluded due to lack of exposure data available.

For HCAI, geolocation was available at the five-digit ZIP level. Daily 100 m pollution estimates were aggregated to ZIP-level daily means via block-group population weighting. For each ED visit, patients were assigned the same-day exposure (lag 0) and lagged exposures for days 1–3 (lag 1–3). For case-crossover analyses, we generated four control periods for each event (1, 2, 3, and 4 weeks prior, matched on weekday) and assigned the corresponding ZIP-level exposures. 1:1, 2:1, 3:1, and 4:1 matching refer to the number of control periods assigned to each case event. Specifically, for each ED or inpatient visit (case day), we selected control days occurring exactly 1, 2, 3, and 4 weeks prior to the event date, matched on day of week to control for weekly temporal patterns. Thus, in a 1:1 design, each case day is matched to one control day (1 week prior); in a 2:1 design, each case day is matched to two control days (1 and 2 weeks prior); and similarly up to four control days (1–4 weeks prior) in the 4:1 specification. The same procedure was applied to inpatient visits.

For CDPH mortality data (2010–2021), residential addresses were geocoded to obtain precise spatial locations. For each decedent, a one-year rolling mean of NO_2 and $PM_{2.5}$ concentrations

prior to the date of death was calculated and assigned as the individual's air pollution exposure. Each death record was matched to one to two living controls selected from the same CDPH dataset, matched on month and year of birth and race-ethnicity to minimize confounding by age and demographic factors. Each death case had T2D as the cause of death, while matched controls drawn from the mortality dataset could include individuals with or without T2D. For these matched controls, the corresponding one-year rolling mean exposures were assigned using the same procedure, based on their geocoded residential addresses and the same temporal exposure windows.

Statistical analysis

Associations between air pollution exposure and multiple health outcomes were evaluated separately for the CHIS, HCAI, and CDPH datasets, using modeling approaches suited to each outcome type and data structure.

To model T2D onset using CHIS data, adult respondents (≥18 years) who reported a diabetes diagnosis were defined as cases, with the self-reported age at diagnosis used to estimate the diagnosis year. Respondents without diabetes served as potential controls. To approximate a population-based risk set, control observations were expanded across years from survey year back to 1990, maintaining age consistency. Up to two controls per case were matched on age, sex, and race-ethnicity to form matched sets. Five-year calendar bins were defined to evaluate potential temporal trends in associations. Logistic regression models were fitted to estimate the effect of air pollution exposure on the odds of incident T2D within each matched set, accounting for matching factors. Covariates included age at diagnosis, sex, race-ethnicity, smoking status, BMI, English proficiency, and additional socioeconomic indicators. Although sex and race-ethnicity were used as matching variables, they were additionally included as covariates in the regression models because matching was not exact for all cases (i.e., some cases had zero, one, or two matched controls). Including these variables as covariates helps account for residual confounding due to incomplete matching and preserves adjustment for these factors across all observations.

Models were fitted separately for each lag (0-3 years) and each 5-year period, as well as for the entire period combined. Effect estimates were expressed as odds ratios (ORs) with 95% confidence intervals (CIs), reflecting the relative odds of T2D per interquartile range (IQR) increase in air pollution exposure. Small sample bins (<30 observations) were excluded to ensure stable estimates.

To model T2D medication use using CHIS data, respondents (≥18 years) reporting current use of any diabetes medication were defined as cases (event = 1), while respondents not reporting medication use served as controls (event = 0). Lagged exposures were calculated for 0–

3 years prior to the survey year to assess potential cumulative effects. Conditional logistic regression models were fitted to estimate the association between air pollution exposure and the odds of diabetes medication use, stratified by survey year to account for temporal clustering. Models incorporated CHIS survey weights to ensure population-representative inference and adjusted for age, sex, race-ethnicity, smoking status, BMI, English proficiency, and other relevant socioeconomic indicators. The person-level weights account for the complex survey design, including unequal selection probabilities, nonresponse, and post-stratification to statewide demographic benchmarks. These weights were incorporated directly into the regression models so that estimated associations reflect population-level effects rather than sample-specific patterns, with variance estimates appropriately accounting for the survey design. Separate models were fitted for each lag period, and effect estimates were expressed as ORs with 95% CIs per IQR increase in air pollution exposure. In this analysis, medication use was defined as a binary outcome indicating whether a survey participant reported using diabetes medication during the survey year, regardless of prior duration of use. The exposure window was therefore aligned with the survey year to capture contemporaneous associations between air pollution exposure and active medication use, rather than medication initiation or cumulative treatment history. This approach is consistent with the structure and limitations of the CHIS data, which do not provide detailed information on medication start dates or duration of use.

For diabetes-related ED visits and hospitalizations in HCAI (2010–2019), we implemented a time-stratified case-crossover design to examine short-term associations with ambient air pollution. This design compares each patient's exposure on the day of the ED visit or hospital admission (case period) with exposures on multiple control days within the same individual, effectively controlling for time-invariant confounders such as sex, race-ethnicity, genetic susceptibility, and underlying comorbidities. For each ED visit, control periods were selected at 1, 2, 3, and 4 weeks prior to the visit, matched on the same day of the week, thereby controlling for day-of-week effects, seasonal trends, and long-term temporal confounding. Conditional logistic regression models were fitted with the case day as the event period and the matched prior days as control periods. Pollutant exposures were scaled by their IQR to standardize effect estimates. Models evaluated same-day exposure (lag 0) and lagged exposures up to three days (lags 1–3) to capture both immediate and delayed effects. Analyses were conducted overall (20% samples) and stratified by race-ethnicity (100% data), allowing assessment of potential disparities in pollutant effects among White, Black, Hispanic, Asian, and Other groups. The 20% sample was selected using a fixed random seed to ensure reproducibility. Given the extremely large resulting sample size, results were stable, and sensitivity checks using alternative seeds yielded nearly identical estimates, indicating robustness to the sampling procedure. ORs and 95% confidence intervals were estimated per IQR increase in pollutant concentration.

For inpatient admissions, length of stay (LOS) was modeled as a continuous outcome to evaluate whether short-term air pollution exposure influenced hospitalization duration. Due to

the right-skewed nature of LOS, generalized linear models (GLMs) with a gamma distribution and log link function were applied. This modeling framework estimates the relative change or mean ratio in LOS associated with air pollution exposure, rather than treating LOS as a binary outcome or as a count of discrete events. The log link was used to accommodate the right-skewed distribution of LOS and to provide interpretable multiplicative effects on the mean length of stay. Exposure metrics included same-day (lag 0) and lagged (lags 1–3) pollutant concentrations to capture immediate and delayed effects. Models were adjusted for demographic factors (age, sex, race-ethnicity, language, and insurance type), ZIP code-level socioeconomic characteristics (e.g., unemployment rate, educational attainment, median household income, marital status), and meteorological variables (e.g., temperature, relative humidity and precipitation). All covariates were included simultaneously within each model to achieve rigorous confounding control, consistent with standard epidemiologic practice for health outcomes such as length of stay. Demographic, socioeconomic, and meteorological variables represent conceptually distinct domains and were retained a priori based on established associations with both air pollution exposure and diabetes-related outcomes. Given the large sample size and the use of GLMs with appropriate distributional assumptions, simultaneous adjustment did not materially inflate variance or compromise model interpretability.

For diabetes-related mortality using the CDPH data (2010–2021), race-ethnicity was reclassified into five categories (Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian, Hispanic, and Other). Records with less than one year of residence in the county were excluded to minimize exposure misclassification. Although all records originated from the CDPH mortality data, diabetes-related deaths were treated as cases and were matched to controls drawn from the same mortality registry who were still alive at the time of the case's death, with matching based on year and month of birth, sex, and race-ethnicity. Exposures were standardized by their IQR to facilitate interpretation. Logistic regression models were used to estimate the associations between air pollution exposures and the odds of mortality. Specifically, we fitted separate models for $PM_{2.5}$ and NO_2, including covariates for age, sex, race-ethnicity, marital status and education level. Here we examined long-term exposure, defined as the annual (365-day) rolling average concentration of pollutants during the year preceding the date of death.

Modelling incidence of diabetes and diabetes medication use from NO_2, $PM_{2.5}$ and O_3 exposure using CHIS data

The CHIS data were pooled from 2011–2019 (biennial waves) to characterize California adults with T2D (Table 12). The analytic sample represents approximately 28.7 million adults, of whom an estimated 2.30 million (8.02%, SE = 0.0017) reported having diabetes. Age distribution of the diabetes population was concentrated among older adults: 55.9% were aged 35–64 years, 41.5% were 65 years and older, and only 2.6% were aged 18–34 years. Males accounted for 52.9% of the diabetes population and females 47.1%. By race and ethnicity, White adults comprised 41.3% of the diabetes population, followed by Latino/Hispanic (35.7%), Asian/Other

(15.0%), African American (7.4%), and American Indian/Alaska Native (0.6%) groups. Socioeconomic patterns showed that more than 55% had incomes at or above 200% of the federal poverty level (FPL) (24.9% between 200–399% and 30.5% ≥400%). Regarding smoking status, 34.4% of adults with diabetes were former smokers, 10.1% current smokers, and 55.5% never smokers. Among adults with diabetes, 83.7% reported using some form of diabetes medication, while 16.3% reported no medication use. Of those using medication, 24.8% used insulin, 76.6% used medications, and 17.7% reported using both insulin and medications.

Table 12. The descriptive characteristics of CHIS Data (2011-2019 biennial waves)

Category	Subgroup	Overall Population		Diabetes Population	
		N (Weighted)	% (Weighted)	N (Weighted)	% (Weighted)
Age Group	18–34	9,123,602	31.8%	59,991	2.6%
	35–64	14,530,212	50.6%	1,286,966	55.9%
	65+	5,060,915	17.6%	956,558	41.5%
Sex	Female	14,675,518	51.1%	1,085,792	47.1%
	Male	14,039,210	48.9%	1,217,723	52.9%
Race/Ethnicity	African American	1,599,553	5.6%	170,703	7.4%
	American Indian/Alaska Native	134,617	0.5%	14,023	0.6%
	Asian/Other	4,668,553	16.3%	344,988	15.0%
	Latino/Hispanic	12,075,073	42.1%	822,503	35.7%
	White	10,236,932	35.7%	951,298	41.3%
FPL Category	<200%	9,813,701	34.2%	1,026,914	44.6%
	200–399%	7,051,338	24.6%	574,602	24.9%
	≥400%	11,849,688	41.3%	702,000	30.5%
Smoking Status	Current Smoker	3,269,317	11.4%	231,818	10.1%
	Former Smoker	6,281,879	21.9%	792,197	34.4%
	Never Smoker	19,163,532	66.7%	1,279,500	55.5%
Diabetes Prevalence	Mean (Proportion)	—	—	0.0802	(SE=0.0017)
Medication Use	No Medication	—	—	375,011	16.3%
	Yes Medication	—	—	1,928,504	83.7%
Insulin Use	No	—	—	1,731,913	75.2%
	Yes	—	—	571,602	24.8%
Pill Use	No	—	—	539,686	23.4%
	Yes	—	—	1,763,829	76.6%
Both Medications	No	—	—	1,896,588	82.3%
	Yes	—	—	406,927	17.7%

When examining the association between NO_2 exposure and the onset of T2D (Figure 7), we found a statistically significant positive relationship when all years were combined. Across all data (1990–2015), NO_2 exposure, normalized through its IQR, demonstrated a consistent association with increased odds of T2D onset (lag 0 OR = 1.013; 95% CI: 1.008–1.017; $p < 0.001$), with similar magnitudes observed at lags 1–3 years, indicating stable health effects. When stratified by diagnostic period, earlier years (1990–1995 and 1995–2000) showed no significant associations between NO_2 and diabetes onset (ORs ≈ 1.00; 95% CIs including 1.00). However, starting from 2000–2005, the relationship became statistically significant (lag 0 OR = 1.012; 95% CI: 1.003–1.022; $p = 0.011$), and the effect magnitude increased slightly in subsequent years. The strongest associations were observed during 2005–2010 (lag 0 OR = 1.019; 95% CI: 1.009–1.029; $p < 0.001$), suggesting a strengthening of NO_2's impact on diabetes onset over time. The 2010–2015 period also exhibited positive and significant associations (lag 0 OR = 1.0154; 95% CI: 1.0001–1.0309; $p = 0.048$). Overall, the findings indicate a temporally consistent and statistically significant relationship between ambient NO_2 exposure and increased risk of T2D onset, particularly from 2000 onward. The growing strength of association in later years may reflect higher exposure susceptibility, improved case detection, or higher precision of residential address/lower exposure misclassification error.

On impact of $PM_{2.5}$ exposure on onset of T2D, a statistically significant and consistent positive association was observed (Figure 8). Across all years combined (1990–2015), elevated $PM_{2.5}$ concentrations were strongly associated with increased odds of diabetes onset (lag 0 OR = 1.074; 95% CI: 1.061–1.087; $p < 0.001$). Similar magnitudes were observed for lags 1–3 years (OR range: 1.075–1.077), indicating a stable and temporally robust effect. When stratified by diagnostic period, the associations persisted across all time intervals but exhibited a pattern of increasing effect magnitude over time. During 1990–1995, the effect was significant (lag 0 OR = 1.036; 95% CI: 1.001–1.072; $p = 0.045$). The associations became more significant by 1995–2000 (lag 0 OR = 1.081; 95% CI: 1.049–1.114; $p < 0.001$) and remained robust through subsequent periods. Between 2000 and 2005, the OR remained around 1.062 (95% CI: 1.037–1.088; $p < 0.001$), while from 2005 to 2010, the effect slightly increased (lag 0 OR = 1.075; 95% CI: 1.051–1.100; $p < 0.001$). The strongest associations were observed during 2010–2015 (lag 0 OR = 1.103; 95% CI: 1.071–1.136; $p < 0.001$), indicating a continued strengthening of the pollutant's impact on diabetes onset. Overall, $PM_{2.5}$ exposure showed a statistically significant and temporally consistent association with higher odds of T2D onset across all examined time periods. The effect magnitude increased slightly over time, suggesting either heightened population vulnerability, changes in $PM_{2.5}$ composition, or improved detection of diabetes cases in later years. These findings reinforce the causal role of fine particulate air pollution in metabolic disease development.

Exposure to ambient O_3 was also found to be positively associated with the onset of T2D, though the magnitude of association was smaller than those observed for $PM_{2.5}$ and NO_2 (Figure

9). When data from all study years (1990–2015) were combined, elevated O_3 exposure was significantly associated with increased odds of diabetes onset (lag 0 OR = 1.028; 95% CI: 1.017–1.039; $p < 0.001$). Similar effects were observed for lags 1–3 (OR range: 1.026–1.027; all $p < 0.001$), demonstrating a consistent relationship between O_3 exposure and diabetes onset across multiple lag structures. Analyses stratified by diagnosis period revealed that the O_3–diabetes association strengthened over time. In the earliest periods (1990–1995 and 2000–2005), associations were weak and statistically non-significant (e.g., 1990–1995 lag 0 OR = 1.006; 95% CI: 0.973–1.040; $p = 0.72$). By 1995–2000, however, the relationship became significant, with consistent positive associations across all lags (lag 0 OR = 1.033; 95% CI: 1.005–1.062; $p = 0.020$). This pattern continued and strengthened in later years, particularly between 2005 and 2010 (lag 0 OR = 1.027; 95% CI: 1.007–1.049; $p = 0.010$) and peaked during 2010–2015, where the associations were most pronounced (lag 0 OR = 1.060; 95% CI: 1.036–1.085; $p < 0.001$). Overall, these results demonstrate a statistically significant and temporally consistent positive association between ozone exposure and diabetes onset, particularly in more recent years. While the magnitude of the O_3 effect was modest compared to $PM_{2.5}$, the increasing strength and consistency of the relationship over time suggest growing public health relevance and possibly heightened population sensitivity or changes in ozone composition and exposure patterns.

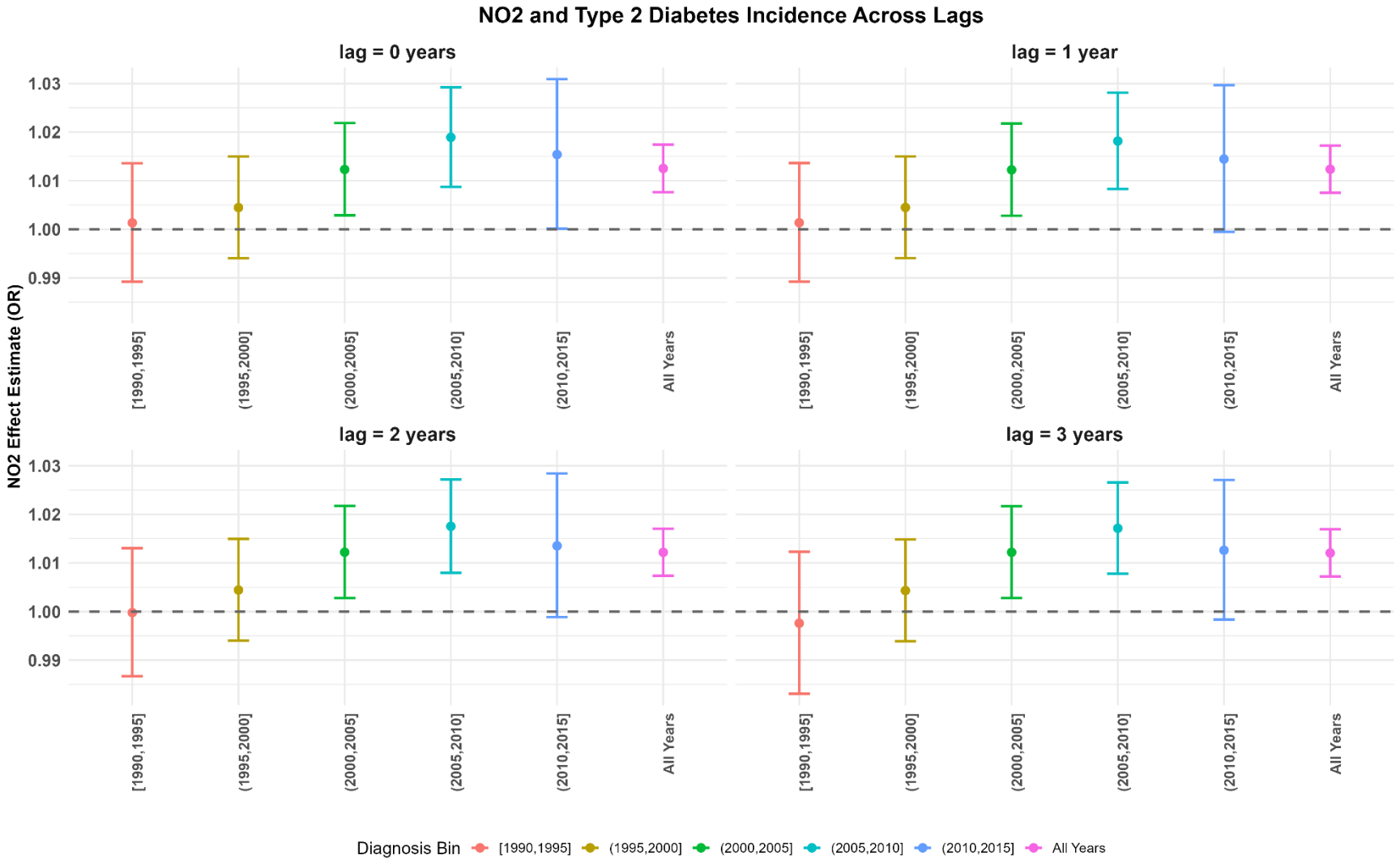


Figure 7. Association of NO_2 with onset of T2D.

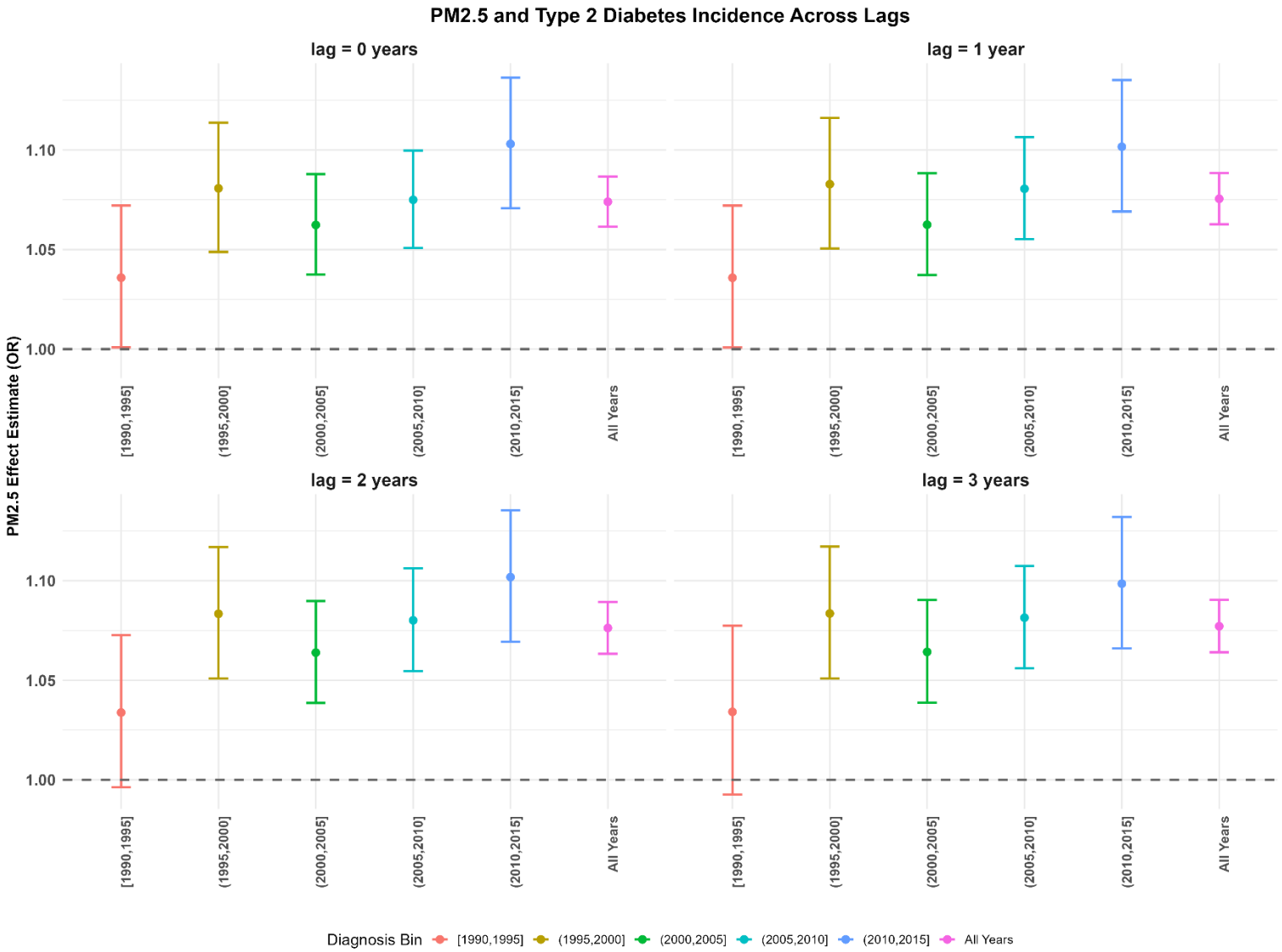


Figure 8. Association of $PM_{2.5}$ with onset of T2D

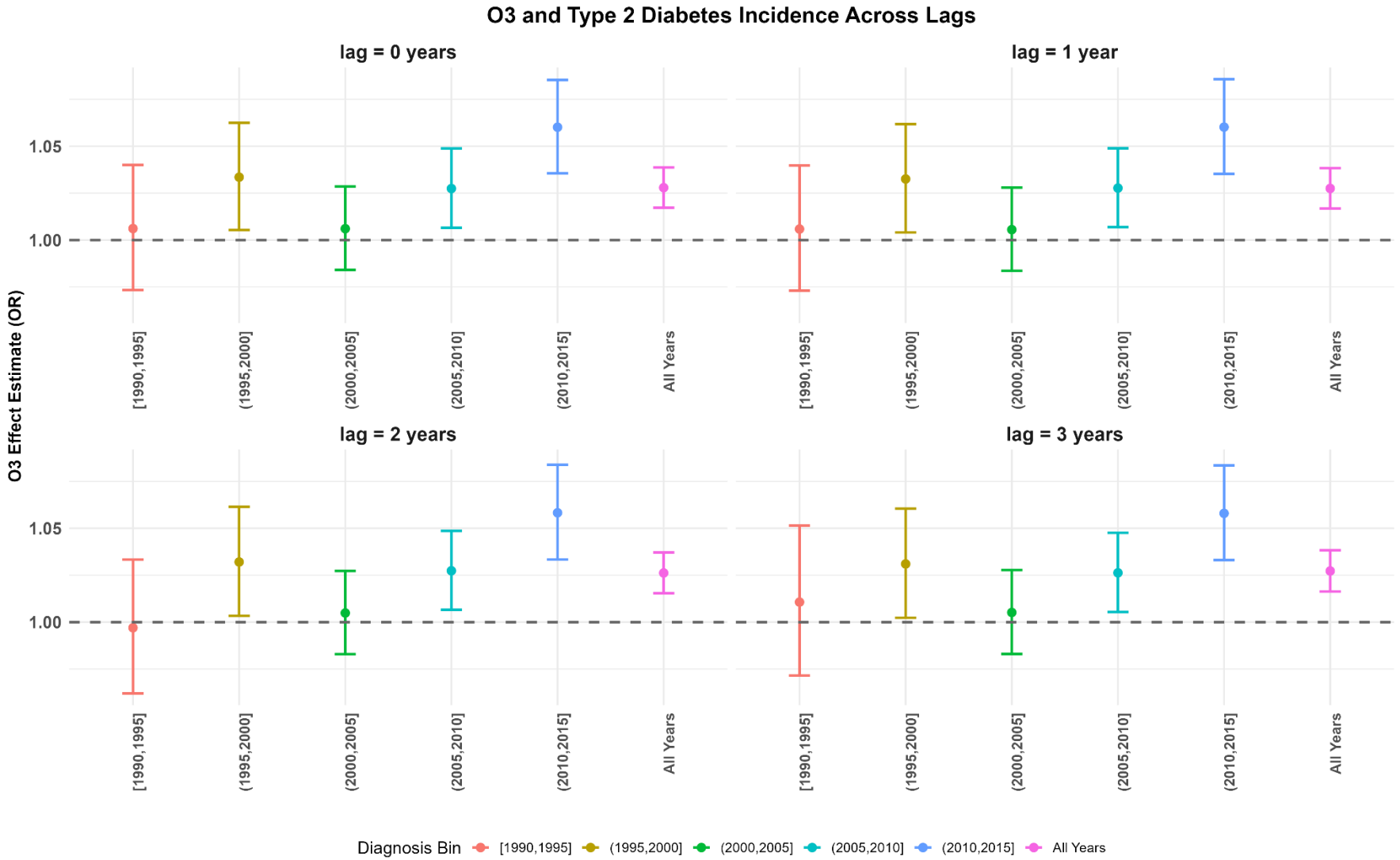


Figure 9. Association of O_3 with onset of T2D

On impact of air pollution exposure on T2D medication use, consistent and statistically significant associations were observed (Figure 10). Across NO_2, $PM_{2.5}$, and O_3, higher pollutant concentrations were linked to increased odds of medication use, suggesting that air pollution may contribute to worsening disease control or increased therapeutic demand among adults with T2D.

For NO_2, the associations were positive but relatively stable across all lag periods, showing little change from lag 0 to lag 3. The estimated odds ratios ranged narrowly from 1.018 (95% CI: 1.014–1.024; $p < 0.001$) at lag 0 to 1.018 (95% CI: 1.013–1.022; $p < 0.001$) at lag 3. These consistent estimates indicate a modest yet persistent relationship between NO_2 exposure and increased medication use. For $PM_{2.5}$, the associations were substantially stronger and exhibited a slight upward trend across the lag structure. The odds ratios increased from 1.070 (95% CI: 1.060–1.081; $p < 0.001$) at lag 0 to 1.073 (95% CI: 1.063–1.084; $p < 0.001$) at lag 3, representing the largest effect magnitude among the three pollutants. This pattern highlights the pronounced impact of fine particulate matter on diabetes management intensity. For O_3, the associations were smaller in magnitude but showed a gradual increase from lag 0 to lag 3. The odds ratios rose from 1.034 (95% CI: 1.026–1.041; $p < 0.001$) to 1.036 (95% CI: 1.028–1.043; $p < 0.001$), suggesting a mild cumulative effect over exposure windows.

Overall, the analyses demonstrate that exposure to ambient air pollutants is associated with greater likelihood of T2D medication use. The magnitude of effect followed the pattern $PM_{2.5} > O_3 > NO_2$, with $PM_{2.5}$ showing the strongest and most consistent associations, while NO_2 exhibited smaller but stable effects across all lag days.

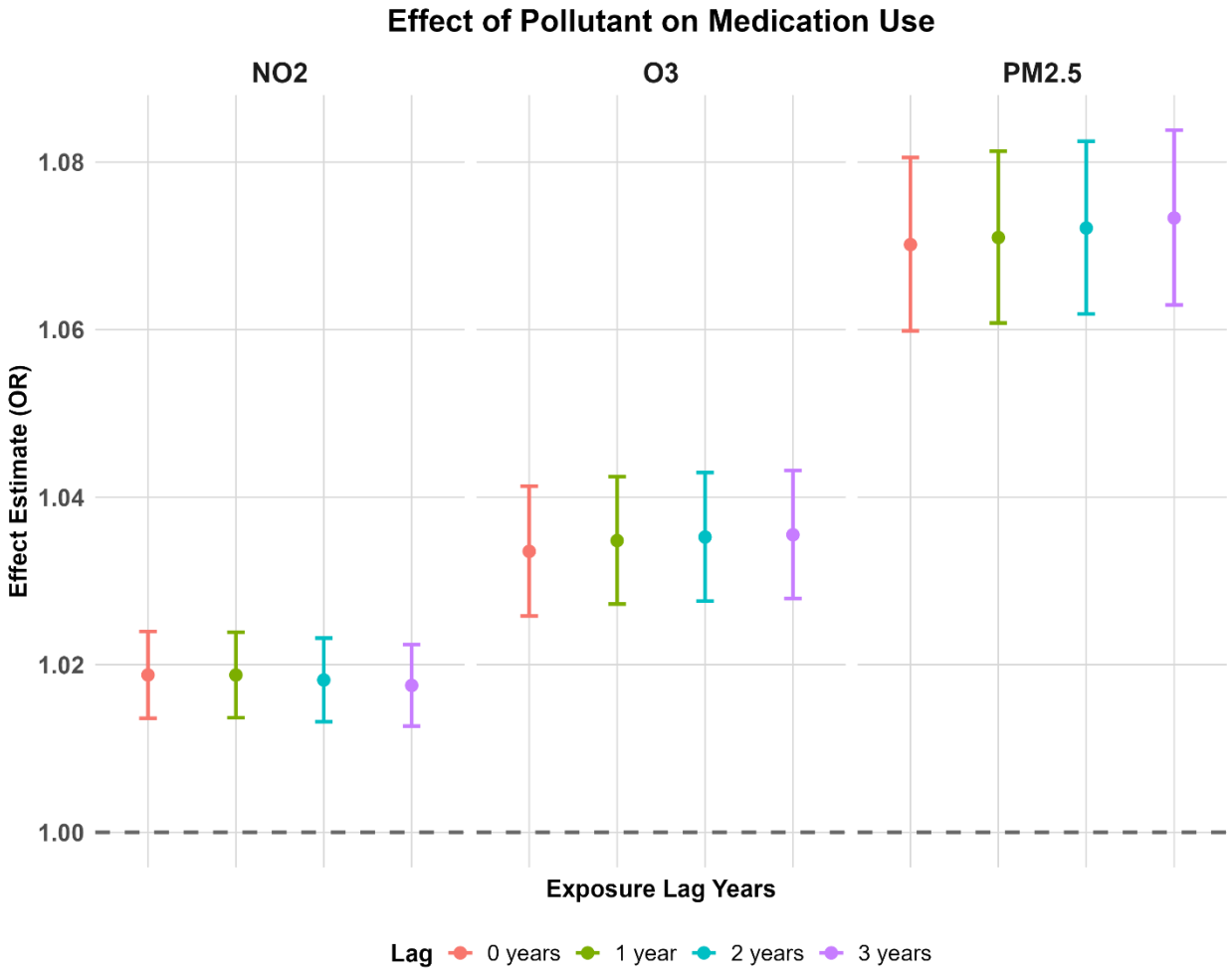


Figure 10. Association of NO_2, $PM_{2.5}$ and O_3 with T2D-related medication use

*Modelling incidence of diabetes and diabetes medication use from **air toxics** exposure using CHIS data*

In the related CARB project 21RD004, we initially modeled ambient concentrations of six air toxics, including benzene, chromium, nickel, lead, 1,3-butadiene, and zinc, at the monthly level. We observed that sample size limitations in certain space-time strata constrained model stability and performance. To address this, we aggregated the data to the annual level, which substantially improved model performance. The resulting models explained between 59% and 90% of the adjusted variance across pollutants. Results from 21RD004 further indicated an inverse association for zinc. Based on this evidence, zinc was excluded from the health outcome analyses in this CARB project (22RD010).

Focusing on the remaining five air toxics, our analysis in this CARB project identified consistent, statistically significant, and positive associations with both diabetes incidence and diabetes medication use across California for all the survey years examined. These findings reinforce the robustness of the observed relationships and support the relevance of these air toxics to population-level metabolic health. Detailed modeling results are presented below, with specific numerical results provided in Supplementary File 4 (Excel).

Air toxics and incidence of T2D

Long-term exposure to benzene was consistently associated with increased incidence of T2D, with effect estimates strengthening markedly over successive diagnosis periods and remaining robust across lag structures (lags 0-3) (Figure 11). For individuals diagnosed between 1990 and 1995, associations were elevate but not statistically significant (e.g., lag 0 OR = 1.11, 95% CI 0.99-1.25). Beginning in the 1995-2000 period, benzene exposure was associated with a clear elevation in diabetes risk (lag 0 OR = 1.28, 1.13-1.44), with similar magnitudes across subsequent lags. Effect sizes increased further in later cohorts, reaching ORs of 1.33 (1.18-1.51) for 2000-2005 and peaking during 2005-2010 (lag 0 OR = 1.60, 1.38-1.86). Elevated risks persisted for diagnoses after 2010. Across all years combined, benzene exposure was associated with a 31-33% higher odds of incident diabetes depending on lag (e.g., lag 0 OR = 1.33, 1.25-1.41), demonstrating a stable and statistically robust relationship between benzene exposure and diabetes onset.

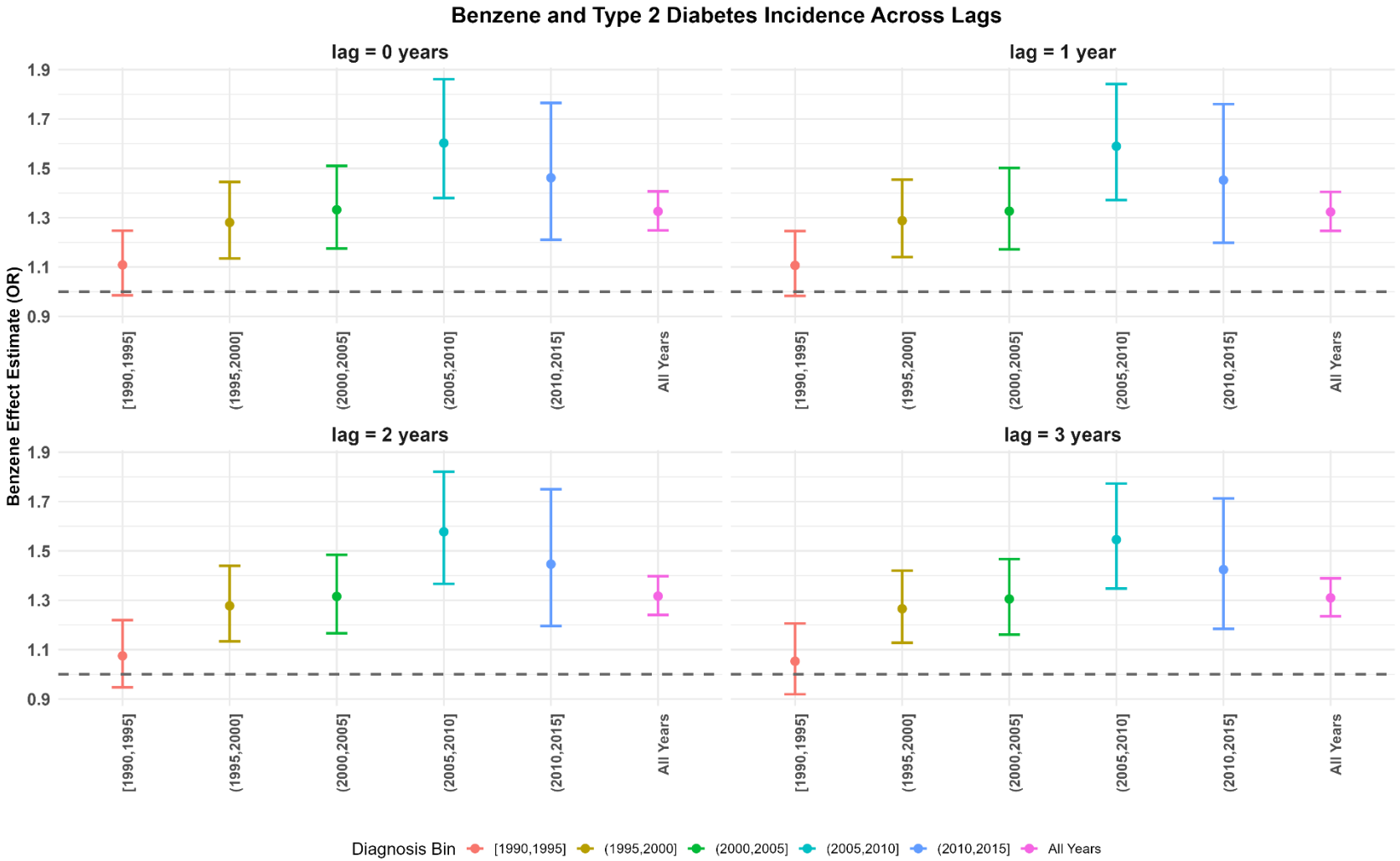


Figure 11. Association of Benzene with onset of T2D

Long-term exposure to chromium was also associated with increased incidence of T2D, with statistically significant associations observed across most diagnosis periods and highly consistent effect estimates across lag structures (lags 0-3) (Figure 12). As with benzene, associations were elevated but not statistically significant among individuals diagnosed between 1990 and 1995 (e.g., lag 0 OR = 1.11, 95% CI 0.98-1.27). Beginning in the 1995-2000 period, chromium exposure was associated with a modest but significant elevation in diabetes incidence (lag 0 OR = 1.16, 1.04-1.30), with nearly identical estimates across all lags. Effect sizes increased during 2000-2005, reaching ORs of approximately 1.29 (95% CI 1.17-1.42), followed by sustained but smaller associations in 2005-2010 (lag 0 OR = 1.17, 1.07-1.28). Associations attenuated and were not statistically significant during 2010-2015. In analyses pooling all years, chromium exposure was associated with an 18% increase in diabetes incidence (lag 0 OR = 1.18, 1.13-1.23), with minimal variation across lag days, indicating a stable short-term exposure-response relationship.

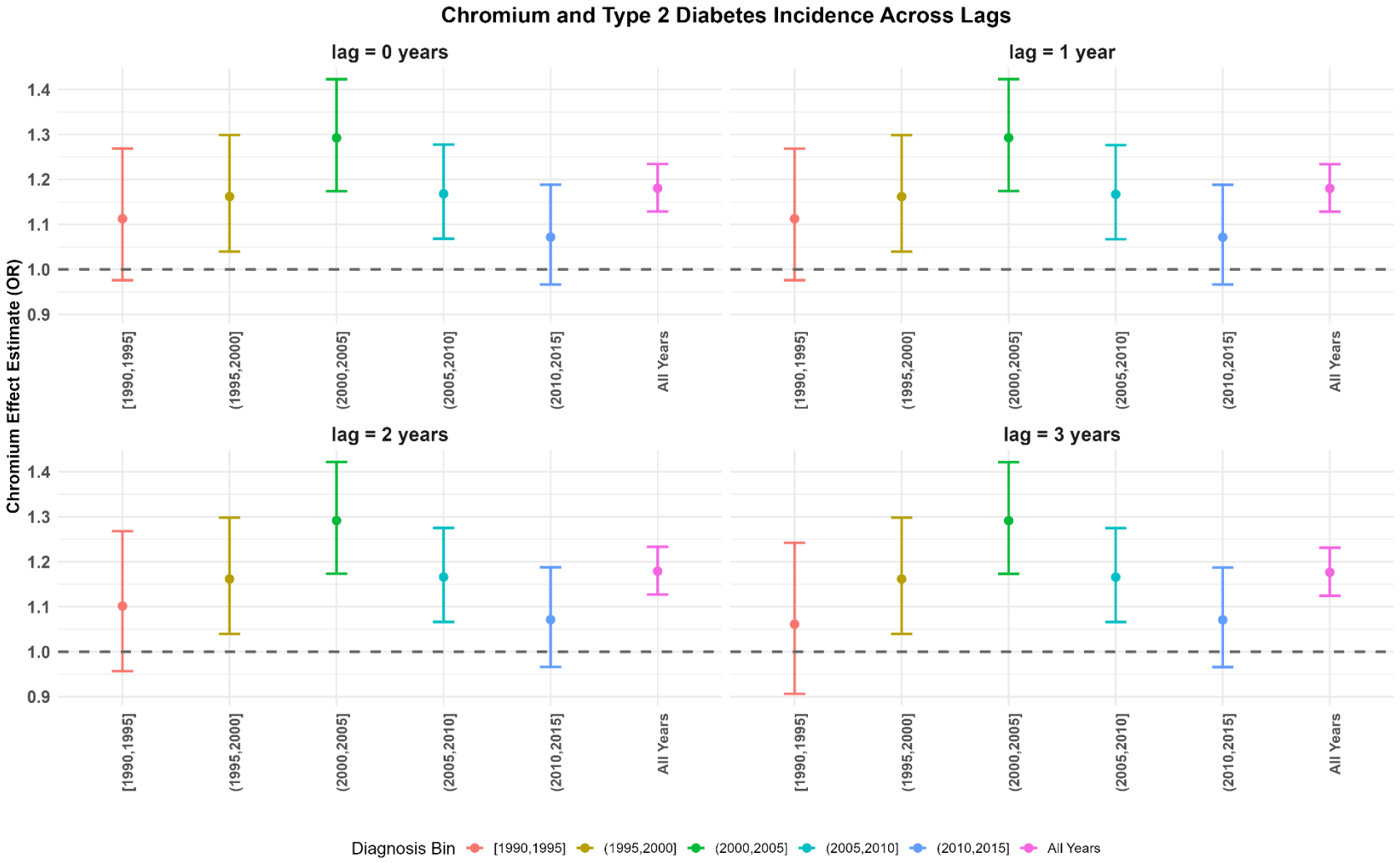


Figure 12. Association of Chromium with onset of T2D

Long-term exposure to nickel was associated with markedly elevated incidence of T2D, with effect sizes substantially larger than those observed for benzene or chromium and strong consistency across lag days (Figure 13). Associations were modest and mostly non-significant in the earliest diagnosis period (1990-1995), although one lag showed borderline significance (lag 2 OR = 1.33, 95% CI 1.01-1.74). From 1995 onward, nickel exposure was robustly associated with increased diabetes incidence across nearly all periods. During 1995-2000, odds ratios exceeded 1.68 across lags, reaching 1.72 at lag 0 (95% CI 1.32-2.25). Effect sizes increased further in 2000-2005 and 2005-2010, with ORs consistently around 1.80-1.85, indicating a strong and stable association during these years. Associations attenuated and were not statistically significant during 2010-2015, suggesting a temporary weakening of the exposure-response relationship. In analyses pooling all years, nickel exposure was associated with approximately a 64-66% increase in diabetes incidence across lags (e.g., lag 0 OR = 1.64, 95% CI 1.45-1.85), underscoring nickel as a particularly potent air toxic in relation to diabetes onset and highlighting its potential importance in the metabolic impacts of ambient air pollution.

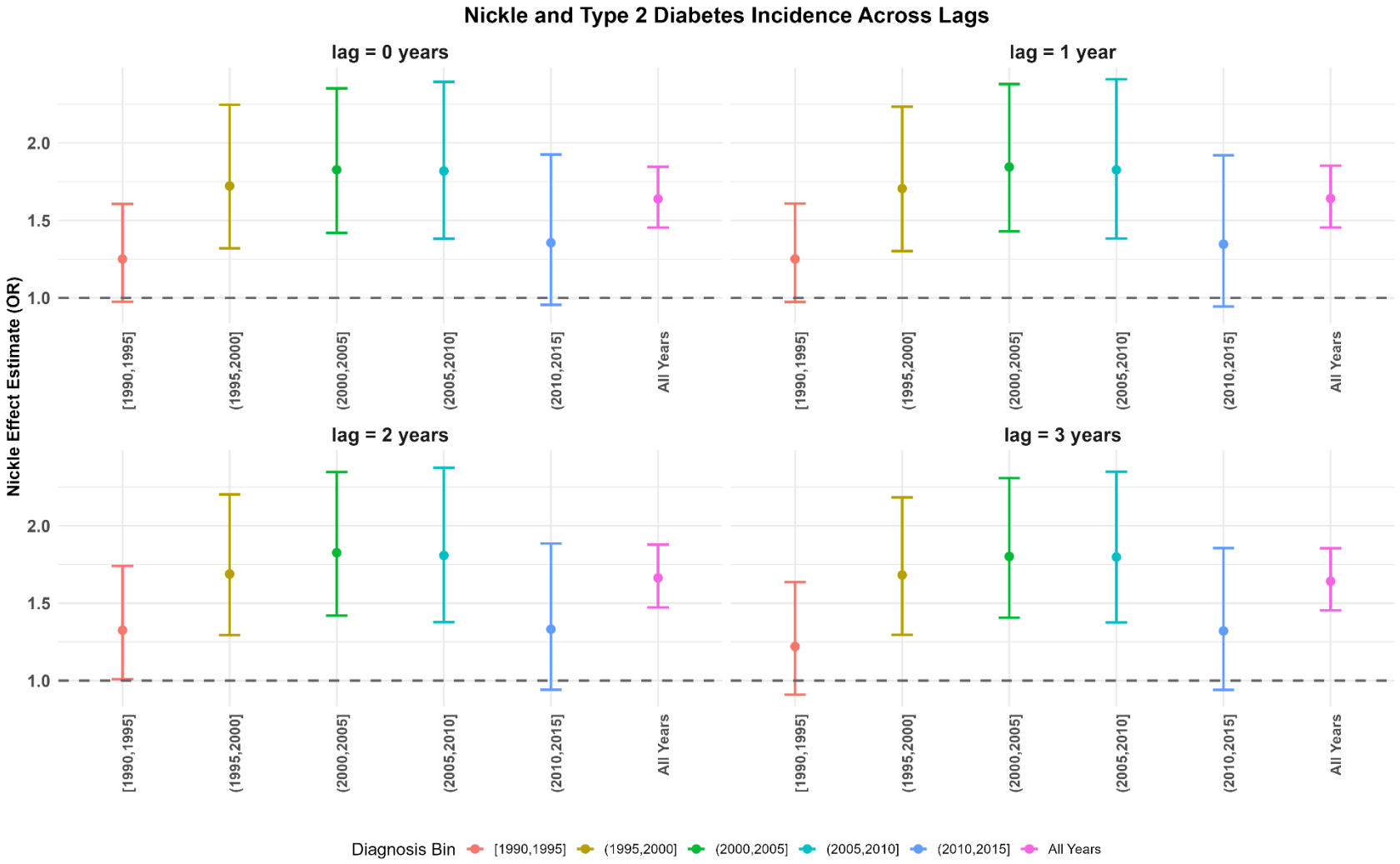


Figure 13. Association of Nickle with onset of T2D

Long-term exposure to lead was associated with a consistently elevated incidence of T2D across most diagnosis periods, with effect sizes that were stable across lag structures, reflecting its chronic rather than acute toxicity profile (Figure 14). In the earliest period (1990-1995), associations were modest but not statistically significant (ORs ranging from approximately 1.04 to 1.06). Beginning in 1995-2000, lead exposure was significantly associated with diabetes incidence, with odds ratios around 1.19 across lags. Stronger and highly consistent associations were observed during 2000-2005 and 2005-2010, with ORs clustered between 1.24 and 1.25, indicating a roughly 24-25% higher risk. The effect estimates attenuated slightly during 2010-2015 (ORs ~1.20-1.21) but remained statistically significant. In analyses pooling all years, long-term lead exposure was significant and associated with an approximately 21% increase in diabetes incidence (ORs ~1.21 across lags), with minimal variation by lag.

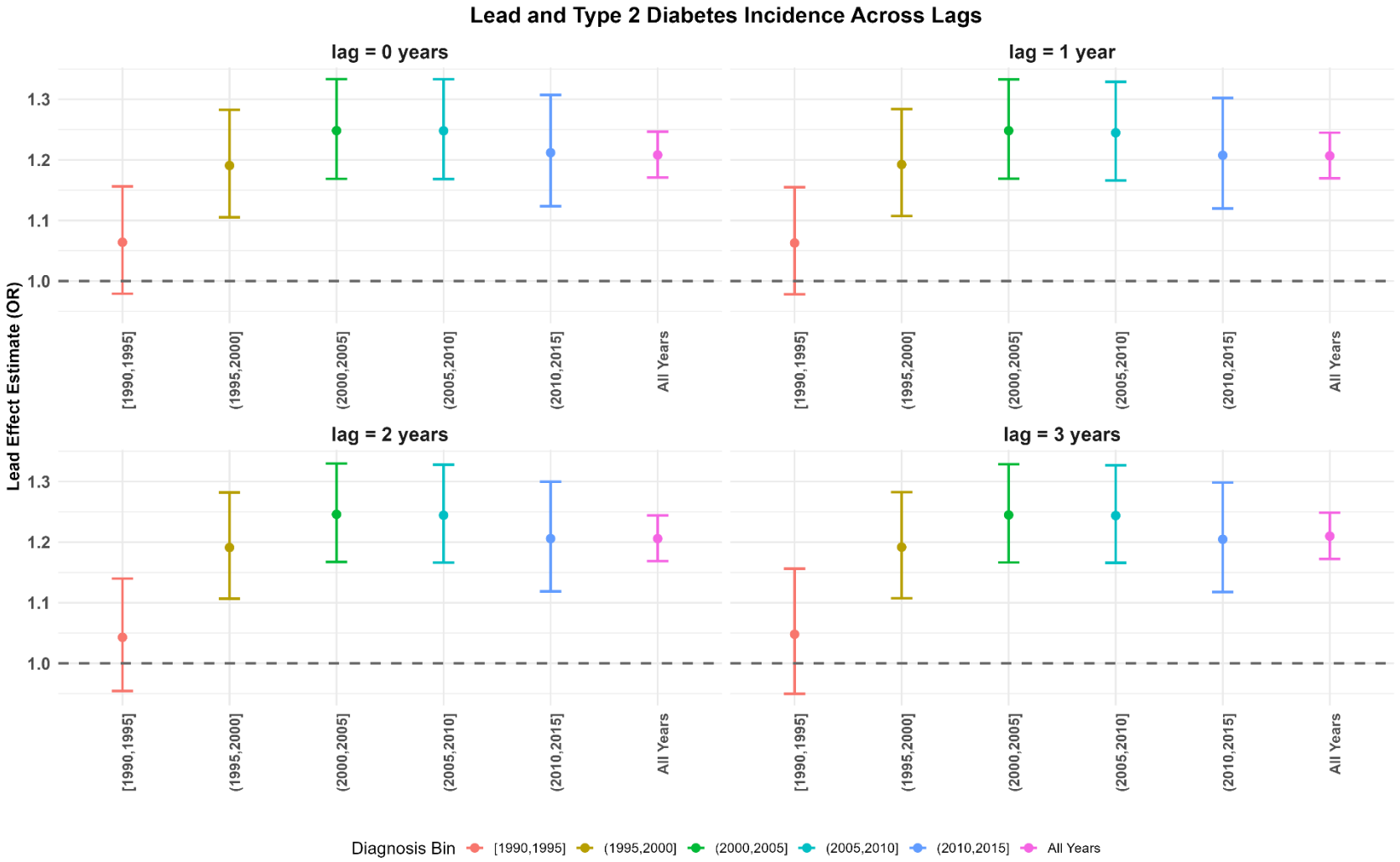


Figure 14. Association of Lead with onset of T2D

Exposure to 1,3-butadiene was associated with elevated incidence of T2D, with effect estimates that varied by calendar period but were generally positive and strengthened in later years (Figure 15). In the earliest period (1990-1995), associations were modest and not statistically significant (ORs approximately 0.93-1.07). During 1995-2000, modest positive associations emerged, with ORs around 1.16-1.19, reaching borderline statistical significance at several lags. Substantially stronger associations were observed in 2000-2005, when ORs ranged from approximately 1.64 to 1.68 across lags, indicating a 60-70% higher incidence. Effects attenuated somewhat in 2005-2010, with mixed significance at shorter lags but increasing ORs at longer lags (up to ~1.38). Associations again strengthened in 2010-2015, with consistently significant ORs of about 1.46-1.52 across lags. In analyses pooling all years, 1,3-butadiene exposure was associated with a 23-30% higher incidence of diabetes (ORs ~1.23-1.30), with slightly larger estimates at longer lags.

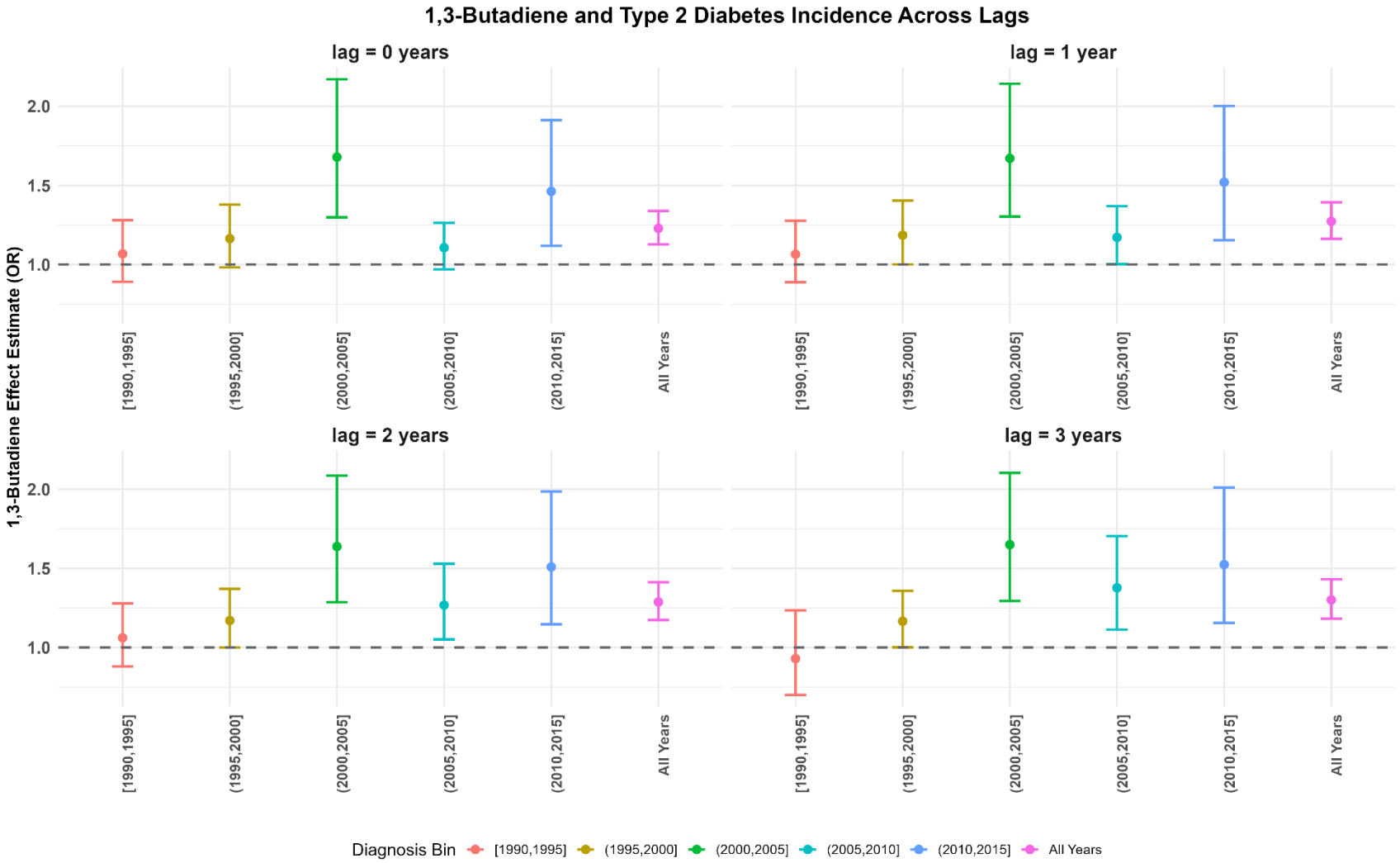


Figure 15. Association of 1,3-Butadiene with onset of T2D

Air toxics and medication use of T2D (Figure 16)

Benzene exposure was positively associated with diabetes medication use in pooled CHIS data from 2011-2019, with consistent and statistically significant effects across all examined lags. At lag 0, benzene exposure was associated with a 25% higher likelihood of medication use (OR = 1.25, 95% CI: 1.19-1.30), with slightly attenuated but still robust associations observed at longer lags. Odds ratios remained elevated at lag 1 (OR = 1.20, 95% CI: 1.16-1.25), lag 2 (OR = 1.19, 95% CI: 1.15-1.23), and lag 3 (OR = 1.19, 95% CI: 1.15-1.23). The monotonic attenuation across lags, coupled with uniformly narrow confidence intervals and highly significant p-values, suggests that benzene exposure is strongly and persistently associated with increased diabetes treatment utilization

Chromium exposure was modestly but consistently associated with increased diabetes medication use. Across all lags, effect estimates were highly stable, indicating a persistent relationship. At lag 0, chromium exposure was associated with a 7% higher likelihood of medication use (OR = 1.07, 95% CI: 1.04-1.11), with nearly identical estimates observed at lag 1 (OR = 1.07, 95% CI: 1.04-1.11), lag 2 (OR = 1.07, 95% CI: 1.04-1.11), and lag 3 (OR = 1.07, 95% CI: 1.04-1.11). The minimal attenuation across lags and the narrow confidence intervals suggest a stable short-term association between chromium exposure and diabetes medication utilization.

Nickel exposure was strongly associated with increased diabetes medication use, with consistently elevated odds across all short-term lags. The largest effect was observed at lag 0, where nickel exposure was associated with a 32% higher likelihood of medication use (OR = 1.32, 95% CI: 1.24-1.40). Elevated associations persisted at lag 1 (OR = 1.24, 95% CI: 1.19-1.30), lag 2 (OR = 1.25, 95% CI: 1.19-1.30), and lag 3 (OR = 1.22, 95% CI: 1.17-1.27), with all estimates highly statistically significant. The modest attenuation across increasing lags suggests a robust and sustained short-term relationship between nickel exposure and intensified diabetes management.

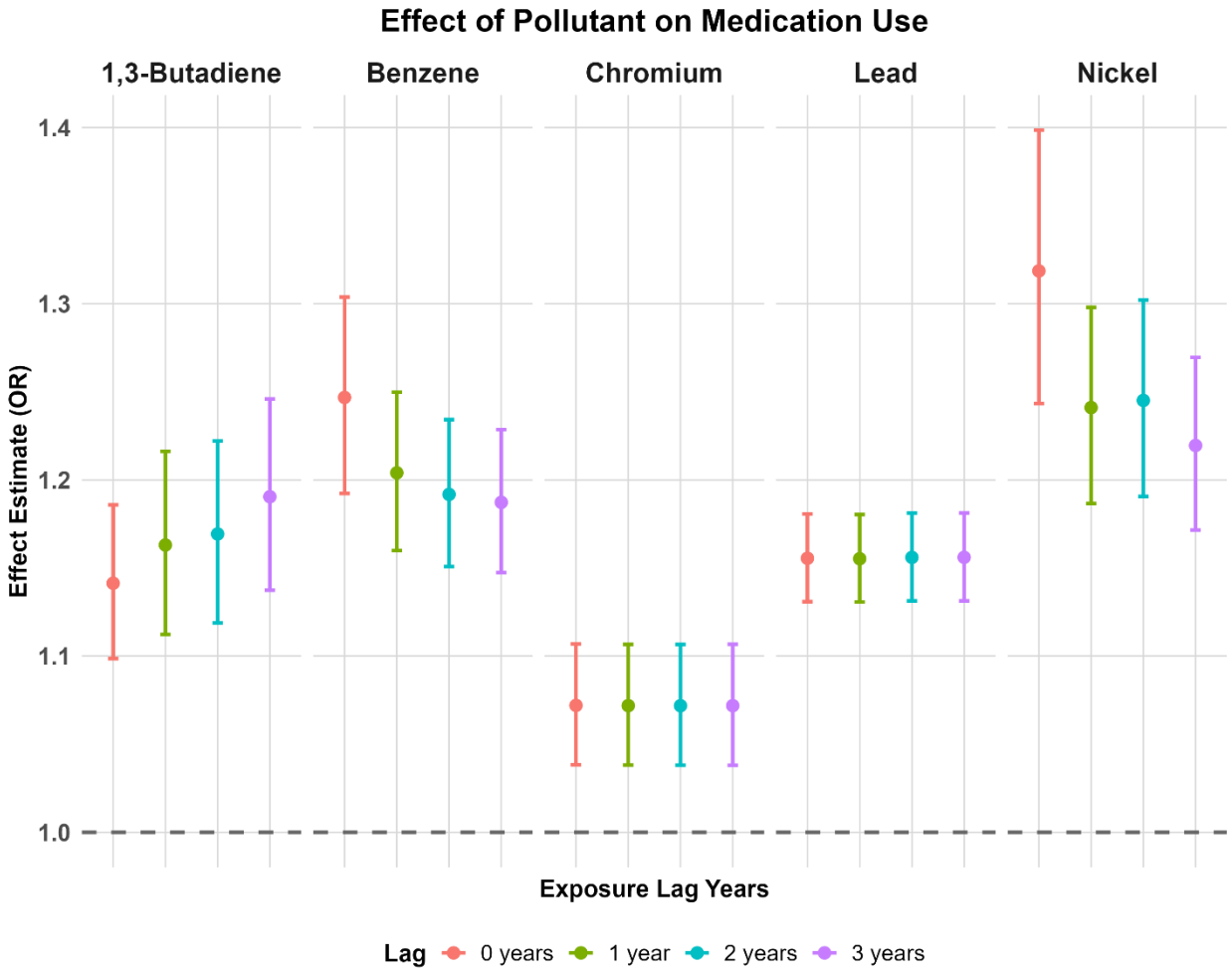


Figure 16. Association of air toxics with medication use across four different lags.

Lead exposure was consistently associated with higher diabetes medication use, with remarkably stable effect estimates across all short-term lags. At lag 0, lead exposure was associated with a 16% increase in the odds of medication use (OR = 1.16, 95% CI: 1.13-1.18), and virtually identical associations were observed at lag 1 (OR = 1.16, 95% CI: 1.13-1.18), lag 2 (OR = 1.16, 95% CI: 1.13-1.18), and lag 3 (OR = 1.16, 95% CI: 1.13-1.18). All associations were highly statistically significant.

Exposure to 1,3-butadiene was also positively associated with diabetes medication use in the pooled CHIS data from 2011-2019, with increasing effect estimates across successive lags. At lag 0, 1,3-butadiene exposure was associated with a 14% higher odds of medication use (OR = 1.14, 95% CI: 1.10-1.19). The association strengthened at lag 1 (OR = 1.16, 95% CI: 1.11-1.22) and lag 2 (OR = 1.17, 95% CI: 1.12-1.22), reaching the largest magnitude at lag 3 (OR = 1.19, 95% CI: 1.14-1.25). All lag-specific estimates were statistically significant.

Modelling diabetes mellitus ED visits from NO2 and PM2.5 exposure using HCAI data

Between 2010 and 2019, the HCAI dataset included 39.3 million total emergency department (ED) visits, of which 7.87 million visits remained after removing accidental causes and duplicates (Table 13). The cleaned dataset consisted of 4,336,098 visits among White patients, 1,472,184 among Hispanic patients, 1,043,753 among Black patients, 548,355 among Asian patients, and 467,991 among patients of Other racial/ethnic backgrounds. This large and demographically diverse sample provides strong statistical power to examine race-ethnicity-specific impacts of air pollution exposure on diabetes-related ED outcomes.

For air pollution, average NO_2 exposures were highest among Hispanic (8.91 ppb) and Black (8.88 ppb) patients, followed by Asian (8.39 ppb), White (8.10 ppb), and Other (8.09 ppb) groups. Similarly, mean $PM_{2.5}$ exposures were slightly higher among Hispanic (8.84 μg m^{-3}) and Black (8.76 μg m^{-3}) patients than among White (8.67 μg m^{-3}), Asian (8.44 μg m^{-3}), and Other (8.42 μg m^{-3}) patients. Both pollutants exhibited substantial variability, with NO_2 interquartile ranges (IQRs) of 7.39–7.83 ppb and $PM_{2.5}$ IQRs of 3.38–3.65 μg m^{-3} across racial/ethnic groups, reflecting exposure disparities and spatial heterogeneity across California communities.

Table 13. Sample size population and exposure statistics for ED visits.

Race Ethnicity	Total ED Visits	Remove Accidents/ Duplicates	Pollut NO_2 (ppb)			Pollut $PM_{2.5}$ (ug m^{-3})		
			Mean	Std	IQR	Mean	Std	IQR
Black	5,218,765	1,043,753	8.88	5.66	7.74	8.76	2.84	3.40
White	21,680,490	4,336,098	8.10	5.48	7.39	8.67	3.04	3.65
Asian	2,741,775	548,355	8.39	5.63	7.83	8.44	2.84	3.53
Hispanic	7,360,920	1,472,184	8.91	5.70	7.79	8.84	2.82	3.38
Other	2,339,955	467,991	8.09	5.47	7.52	8.42	2.95	3.60

Across the study period, NO_2 exposure was associated with consistent, though modest, statistically significant increases in diabetes-related ED visits (Figure 17). For each individual lag, the estimated odds ratios changed very little as the number of matched controls increased from 1:1 to 4:1. For example, at lag 0, the OR ranged only slightly from 1.0072 for 1:1 matching to 1.0065 for 4:1 matching, indicating that the choice of control ratio had minimal impact on effect estimates. Conversely, for each individual control ratio, the estimated effect decreased gradually as the exposure lag increased from 0 to 3 days. For 1:1 matching, the OR declined from 1.0072 at lag 0 to 1.0026 at lag 3, and similar modest declines were observed across other matching ratios. Overall, these patterns suggest that recent NO_2 exposure (lag 0–1) has the strongest association with ED visits, while the influence of control matching strategy is minimal, and longer lags show only slight attenuation of the effect.

Across racial and ethnic groups, NO_2 exposure was still consistently associated with increased odds of diabetes-related ED visits, with some differences in magnitude by group (Figure 18). For each group, increasing the number of matched controls from 1:1 to 4:1 led to a gradual increase in estimated odds ratios, although the effect of additional controls was relatively modest compared with inter-group differences. For example, at lag 0, the OR for Black individuals increased from 1.026 (1:1) to 1.063 (4:1), whereas for Hispanic individuals the corresponding increase was from 1.030 to 1.070, indicating that Hispanic and Other race-ethnicity groups had the largest absolute effects. Across lags within each control ratio, effect estimates generally decreased slightly as lag increased from 0 to 3 days. For instance, for 1:1 matching among Black individuals, the OR declined from 1.026 at lag 0 to 1.009 at lag 3, showing modest attenuation over time. Overall, these patterns suggest that the impact of NO_2 is strongest for recent exposure (lag 0–1), and the magnitude of effect varies by race-ethnicity, with Hispanic, Asian, and Other groups showing the largest associations, while Black and White groups had slightly lower but still significant increases in odds.

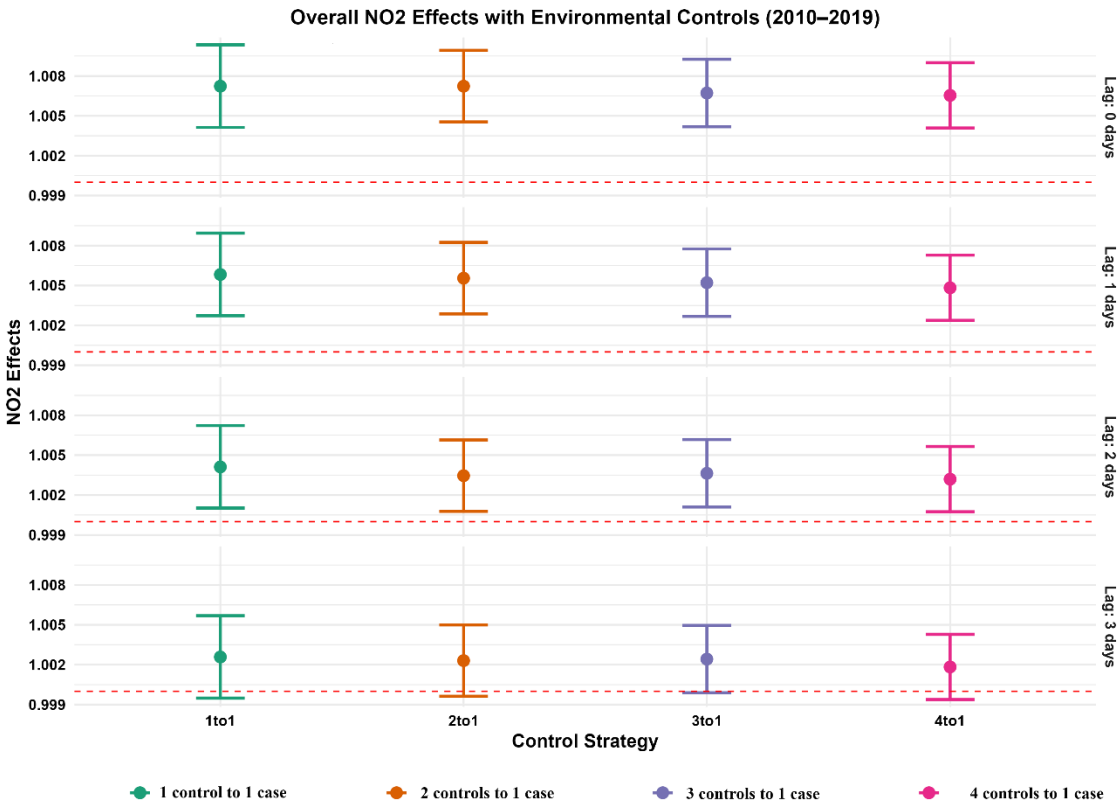


Figure 17. The overall impact of NO_2 exposure on ED visits.

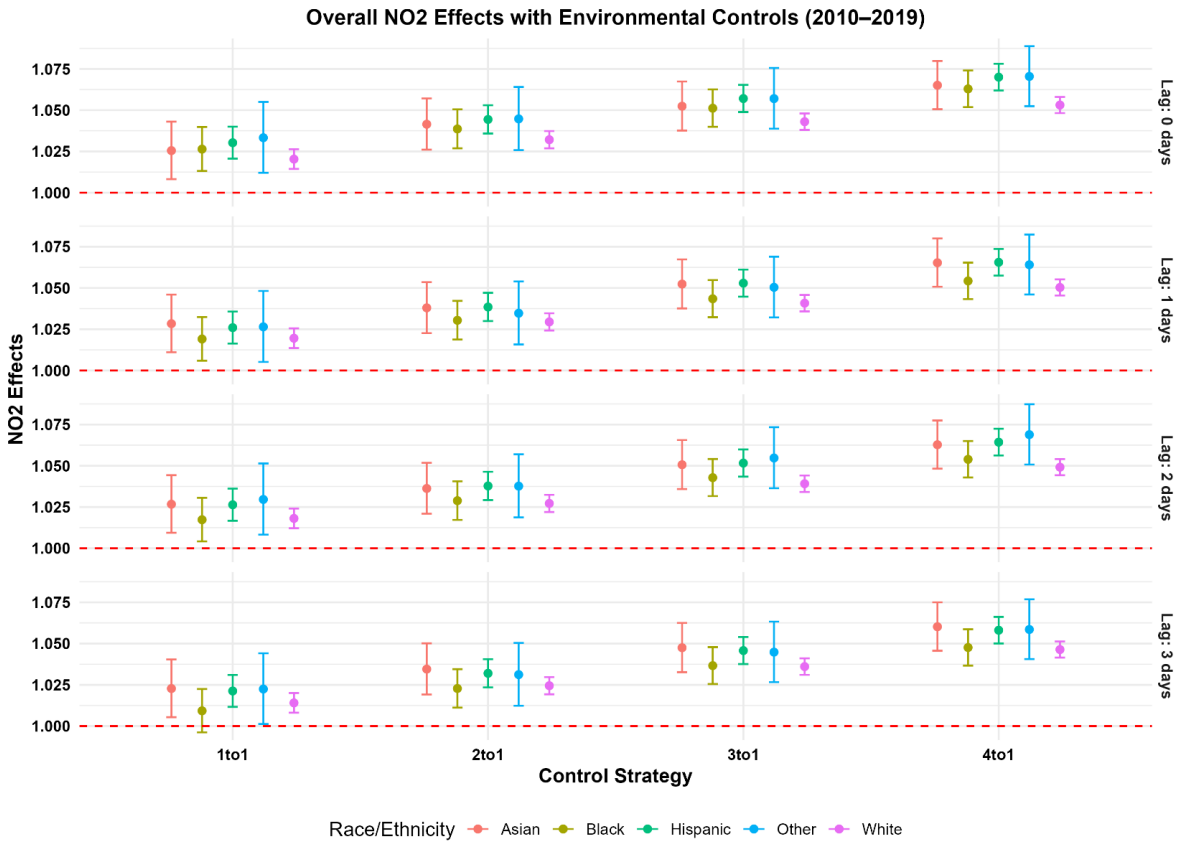


Figure 18. The overall impact of NO_2 exposure on ED visits stratified by race-ethnicity.

Note: Effect estimates are scaled to race-ethnicity and control-group specific NO_2 IQRs to account for heterogeneity in exposure distributions across groups, ensuring that estimates reflect within-group exposure variability and support interpretable comparisons of relative susceptibility within each race-ethnicity. This scaling approach was applied consistently across all models involving race-ethnicity and/or control group specific stratification.

For $PM_{2.5}$ exposure across all adults, it was also found to be associated with statistically significant increased odds of diabetes-related ED visits over the study period (Figure 19). For each lag, increasing the number of matched controls from 1:1 to 4:1 resulted in slight increases in effect estimates, although the magnitude of change was small; for example, at lag 0, the OR increased from 1.004 (1:1) to 1.006 (4:1), indicating that expanding the control set had minimal impact on estimated associations. Across lags within a given control strategy, the effect estimates showed modest attenuation over time: for 1:1 matching, the OR decreased from 1.004 at lag 0 to 1.001 at lag 2 and to 1.001 at lag 3. Overall, $PM_{2.5}$ effects had the strongest associations observed for concurrent exposure (lag 0) and slightly weaker associations for exposures over 1–3 days. These findings suggest that recent $PM_{2.5}$ exposure may contribute to increased diabetes-related ED visit risk.

Race-ethnicity–stratified analyses for $PM_{2.5}$ exposure (Figure 20) also showed consistent increased odds of diabetes-related ED visits across all groups, with the magnitude of effect generally highest for Black and Hispanic adults. Increasing the number of matched controls from 1:1 to 4:1 led to small but consistent increases in estimated effects across all races; for example, at lag 0, Black adults' ORs increased from 1.022 (1:1) to 1.051 (4:1), and Hispanic adults' ORs increased from 1.026 to 1.054. Across lags within a given control strategy, effect estimates typically attenuated slightly over time: for 1:1 matching, Black adults' OR decreased from 1.022 at lag 0 to 1.002 at lag 3, while similar modest declines were observed for White, Asian, Hispanic, and Other adults. Overall, Black and Hispanic adults consistently exhibited the largest associations with $PM_{2.5}$ exposure, suggesting that these populations may experience relatively higher risk of diabetes-related ED visits in response to fine particulate matter. The temporal attenuation across lags and modest effect of control expansion indicate that recent exposure contributes most strongly, with exposures over 1–3 days producing slightly weaker associations.

We also conducted similar analyses for individual years and with various stratification strategies, including health insurance pay type, region, primary language spoken, race-ethnicity, sex and age group. The details of those analyses are in Supplementary File 1.

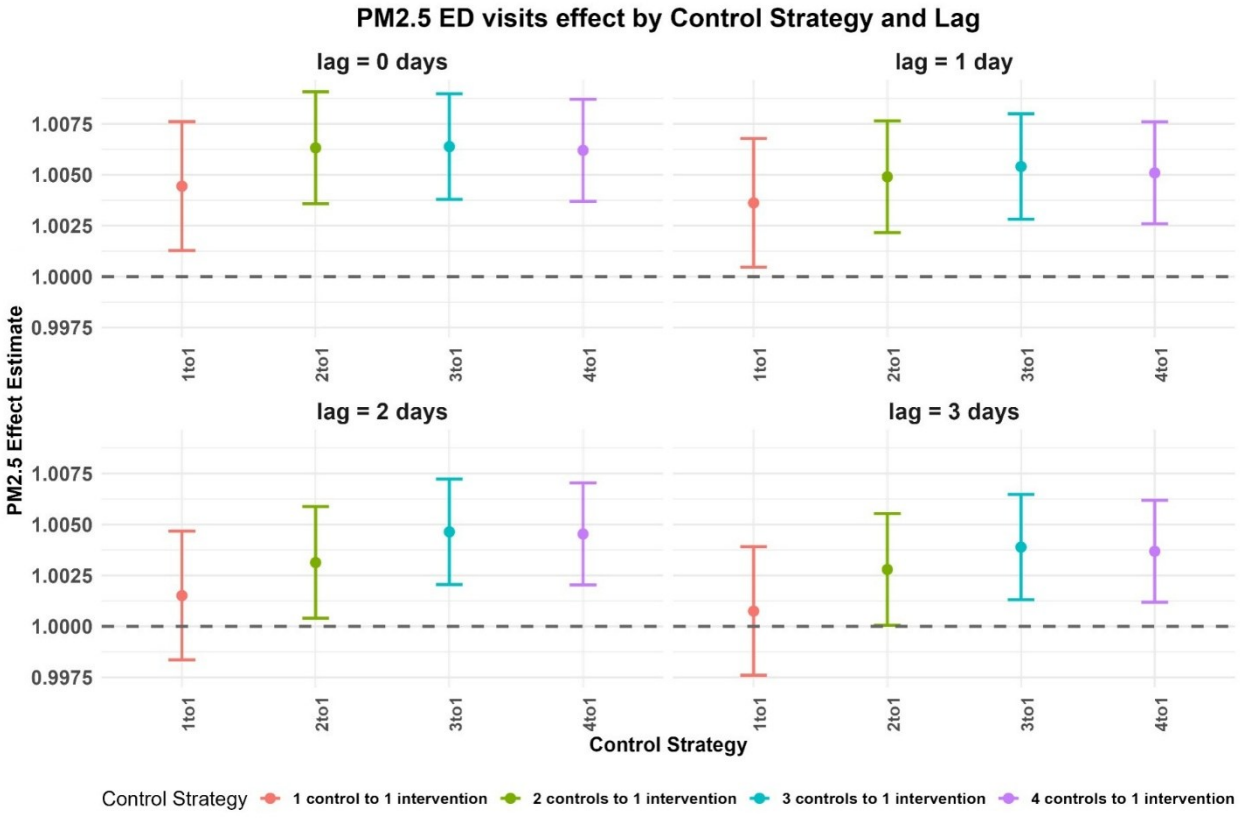


Figure 19. The overall impact of $PM_{2.5}$ exposure on ED visits.

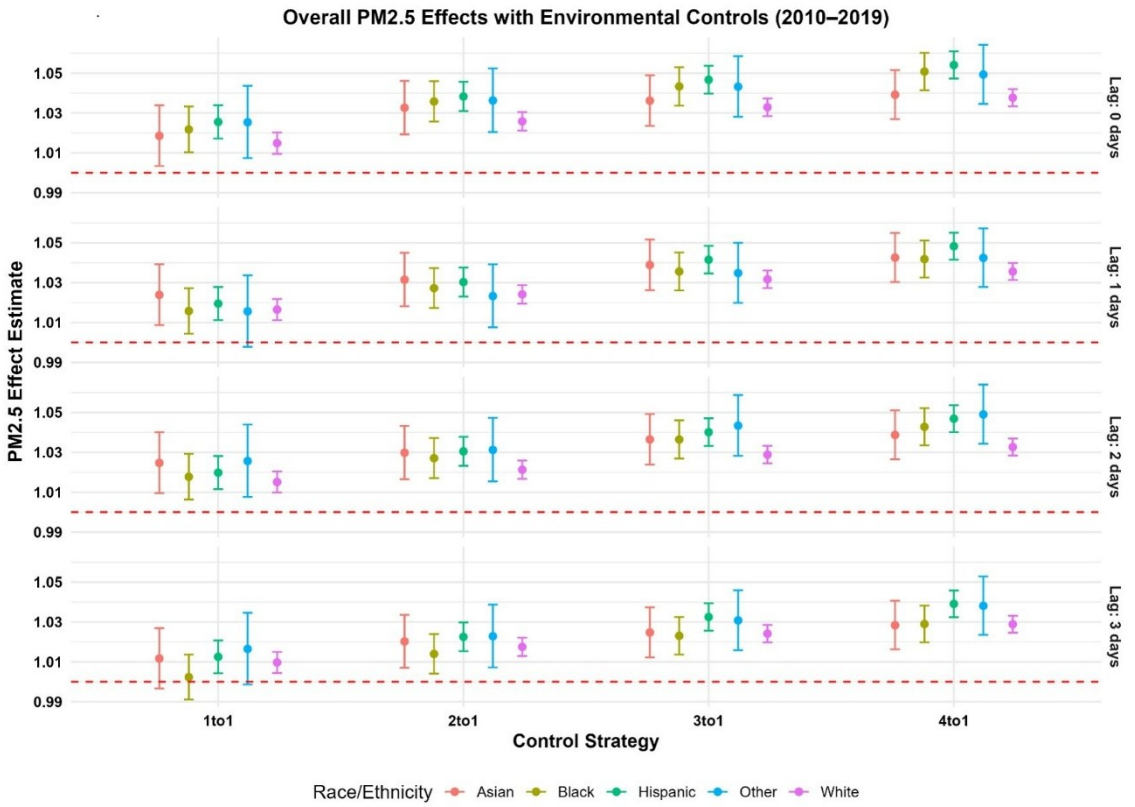


Figure 20. The overall impact of $PM_{2.5}$ exposure on ED visits stratified by race-ethnicity.

Modelling diabetes mellitus hospital admissions from NO_2 and $PM_{2.5}$ exposure using HCAI data

Between 2010 and 2019, a total of approximately 29.07 million inpatient (IP) records were identified in the HCAI dataset across all race and ethnicity groups, of which about 6.01 million (20.7%) remained after removing records related to accidents and duplicate entries (Table 14). The "Other" race-ethnicity category contributed the largest absolute number of hospitalizations (5.38 million), contributed by its broader definition encompassing multi-racial and unspecified individuals. Among other race-ethnicity categories, White patients accounted for the largest share (379,619 visits), followed by Hispanic (112,559), Asian (69,845), and Black (67,871) individuals.

Across racial and ethnic groups, exposure levels showed modest variation. Mean NO_2 concentrations ranged from 8.20 ppb in White to 9.16 ppb in Hispanic patients, with IQRs between 7.5 and 8.2 ppb, suggesting moderate within-group variability. Similarly, mean $PM_{2.5}$ levels varied narrowly between 8.40 $\mu g\ m^{-3}$ (Other) and 8.85 $\mu g\ m^{-3}$ (Hispanic), with IQRs from 3.3 to 3.7 $\mu g\ m^{-3}$. Overall, both NO_2 and $PM_{2.5}$ exposures exhibited slightly higher averages

among Hispanic and Black populations, reflecting potential spatial overlap between higher pollution burdens and these demographic communities.

Across the entire 2010–2019 study period, short-term exposure to NO_2 was significantly associated with increased odds of hospital admissions for diabetes-related conditions (Figure 21). Conditional logistic regression models using daily data demonstrated a consistent and statistically significant positive relationship across all lag structures (0–3 days) and control strategies. Overall, each IQR increase in daily NO_2 concentration was associated with a 1–5% higher odds of hospitalization, supporting a robust short-term causal impact of NO_2 on acute metabolic health outcomes.

When examined by control ratio, effect estimates showed a modest upward trend as the number of controls per case increased, indicating improved model precision with expanded control sampling. At lag 0, the OR increased from 1.018 (95% CI: 1.006–1.030) under the 1:1 scheme to 1.053 (1.043–1.063) for the 4:1 scheme. A similar pattern was observed for lag 1 (1.016–1.051) and lag 2 (1.011–1.046), while lag 3 estimates remained positive (1.013–1.044) though slightly attenuated. Across lags, the strongest associations were observed at lag 0 and lag 1, suggesting that the adverse effects of NO_2 exposure on hospitalization risk occur within 24–48 hours of exposure.

Across all race-ethnicity groups, short-term exposure to NO_2 was significantly associated with increased odds of diabetes-related hospital admissions, indicating a robust and consistent adverse effect of daily NO_2 exposure (Figure 22). The association remained statistically significant across all control strategies and lags (0–3 days), confirming a causal temporal relationship between acute NO_2 exposure and hospitalization risk. Within each race-ethnicity group, the estimated ORs followed a consistent pattern, with the strongest effects observed at lag 0–1 day, suggesting that NO_2 exposure exerts the greatest influence within 24–48 hours prior to admission. Effect sizes increased modestly with more extensive control sampling (from 1:1 to 4:1), indicating greater stability and precision of the estimated effects. For example, among White patients, ORs ranged from 1.019 (95% CI: 1.012–1.026) under 1:1 matching at lag 0 to 1.048 (1.043–1.054) under 4:1 matching. Similar progressive increases were observed among Hispanic (1.023–1.061) and Asian (1.030–1.070) patients, reflecting consistent exposure-response relationships across model specifications. When comparing across race-ethnicity categories, Asians and individuals classified as “Other” exhibited the largest effect estimates, with ORs exceeding 1.06 under the 4:1 lag 0 configuration, suggesting heightened susceptibility or greater exposure gradients in these populations. Hispanic and White populations showed slightly lower but still statistically significant effects, with ORs around 1.05−1.06 for lag 0−1. In contrast, Black patients demonstrated smaller effect magnitudes (OR ≈ 1.02−1.05) and a slightly slower decline across lags, though still maintaining positive associations.

Across all groups, effect estimates declined modestly with increasing lag days, consistent with the transient nature of air pollution impacts on acute hospital utilization. The persistence of significant associations through lag 3 in most groups supports a cumulative short-term exposure effect. Overall, the results demonstrate a statistically significant and temporally coherent causal relationship between daily NO_2 exposure and diabetes-related hospital admissions, with notable variability in effect magnitude by race-ethnicity, potentially reflecting differential vulnerability, exposure patterns, or contextual factors.

For $PM_{2.5}$, across the 2010–2019 study period, higher daily exposure was consistently and significantly associated with increased odds of diabetes-related hospital admissions (Figure 23). The associations were statistically robust across all matching control strategies and lag periods (0–3 days), indicating a persistent adverse impact of $PM_{2.5}$ exposure on hospitalization risk. At lag 0, which represents exposure on the same day of admission, the estimated odds ratios increased gradually with broader control sampling—from 1.02 (95% CI: 1.01–1.03) under the 1:1 control ratio to 1.04 (95% CI: 1.04–1.05) under 4:1 control. Similar incremental patterns were observed for lags 1 through 3, confirming that results were consistent and not sensitive to the control selection strategy. In terms of temporal trends, the strongest effects were generally observed for lag 0 and lag 1 in a way like those of NO_2, indicating that $PM_{2.5}$ exposure on the day of or one day prior to admission exerts the most immediate influence on hospital utilization. The magnitude of association diminished slightly over lags 2–3 but remained statistically significant through lag 3, supporting a short-term cumulative exposure effect.

For $PM_{2.5}$ impact on hospital admissions across race–ethnicity groups, short-term increases in $PM_{2.5}$ concentration were consistently associated with a statistically significant increase in the odds of hospital admissions across all racial and ethnic categories (Figure 24). The associations remained robust across all lag structures (lags 0–3 days) and control strategies, underscoring a strong and stable causal relationship between daily $PM_{2.5}$ exposure and elevated hospitalization risks. The magnitude of effect was modest but persistent, with ORs generally increasing with broader control strategies (from 1:1 to 4:1). Among racial and ethnic groups, Asians exhibited the strongest and most consistent associations. For lag 0, the ORs increased from 1.033 (95% CI: 1.017–1.049) under the 1:1 control strategy to 1.064 (95% CI: 1.051–1.077) under the 4:1 control strategy. Similar elevated risks persisted across lags 1–3, with ORs typically exceeding 1.03, indicating heightened sensitivity of this group to fine particulate pollution. Black populations also demonstrated pronounced and statistically significant associations at early lags, with ORs at lag 0 ranging from 1.030 (95% CI: 1.015–1.046) to 1.060 (95% CI: 1.048–1.072) from matching 1:1 to 4:1, remaining elevated through lag 1 before gradually attenuating by lag 3. For Hispanic and White populations, the associations were similarly strong and consistent across all lags. Among Hispanics, lag 0 ORs rose from 1.019 (95% CI: 1.008–1.031) to 1.048 (95% CI: 1.038–1.057) from matching 1:1 to 4:1 under broader matching strategies, with statistically significant elevations persisting through lag 3. Whites showed comparably stable effects, with

lag 0 ORs ranging from 1.015 (95% CI: 1.009–1.021) to 1.037 (95% CI: 1.032–1.043) from matching 1:1 to 4:1. The Other category displayed significant but slightly smaller magnitudes, with ORs at lag 0 increasing from 1.028 (95% CI: 1.008–1.048) to 1.052 (95% CI: 1.035–1.069) from matching 1:1 to 4:1, again showing stability across lags 0–3.

In summary, $PM_{2.5}$ exposure was a statistically significant and consistent predictor of increased hospital admissions across all examined racial and ethnic groups. The effects were strongest and most persistent among Asian and Black populations. These findings emphasize the widespread health burden of $PM_{2.5}$ pollution and its disproportionate impacts across demographic subgroups.

Table 14. Sample size population and exposure statistics for inpatient visits.

Race Ethnicity	Total IP Visits	Remove Accidents/ Duplicates	Pollut NO_2 (ppb)			Pollut $PM_{2.5}$ (ug m^{-3})		
			Mean	Std	IQR	Mean	Std	IQR
Black	339,355	67,871	9.15	5.73	7.95	8.74	2.79	3.34
White	1,898,095	379,619	8.20	5.50	7.53	8.58	3.05	3.69
Asian	349,225	69,845	8.96	5.82	8.15	8.57	2.78	3.45
Hispanic	562,795	112,559	9.16	5.80	7.98	8.85	2.79	3.36
Other	26,919,165	5,379,185	8.21	5.58	7.75	8.40	2.94	3.66

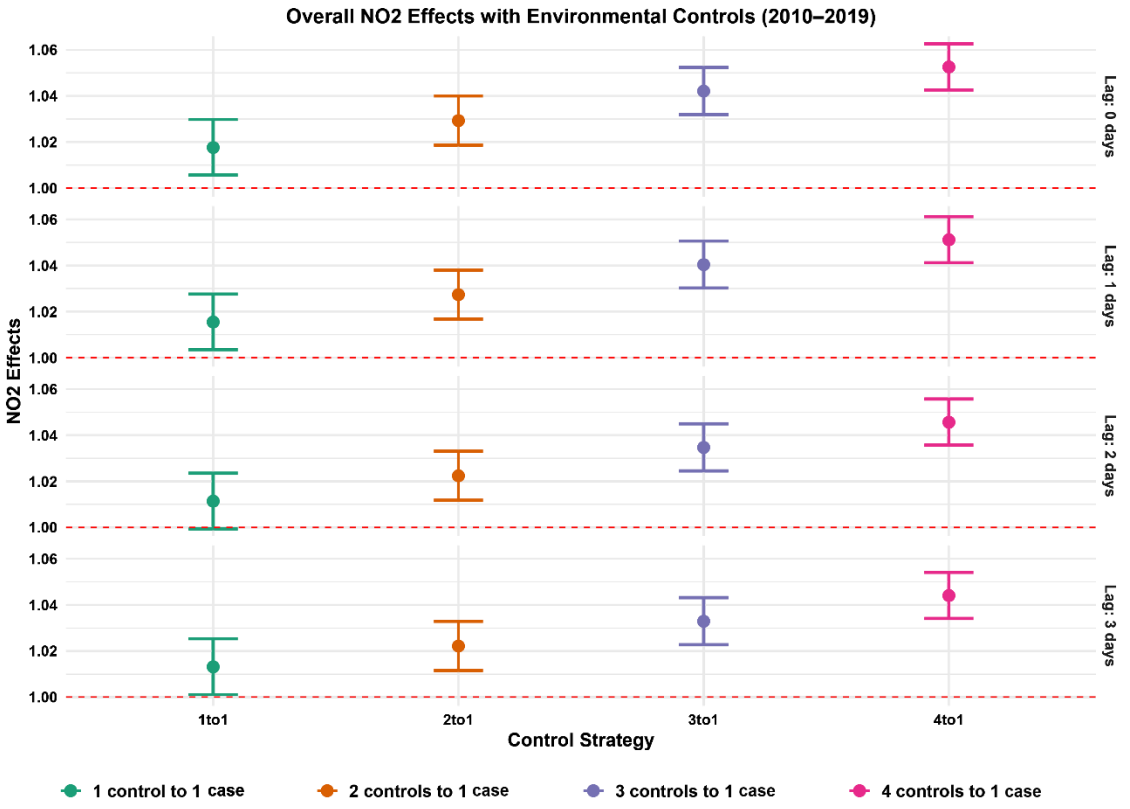


Figure 21. The overall impact of NO_2 exposure on hospital admissions.

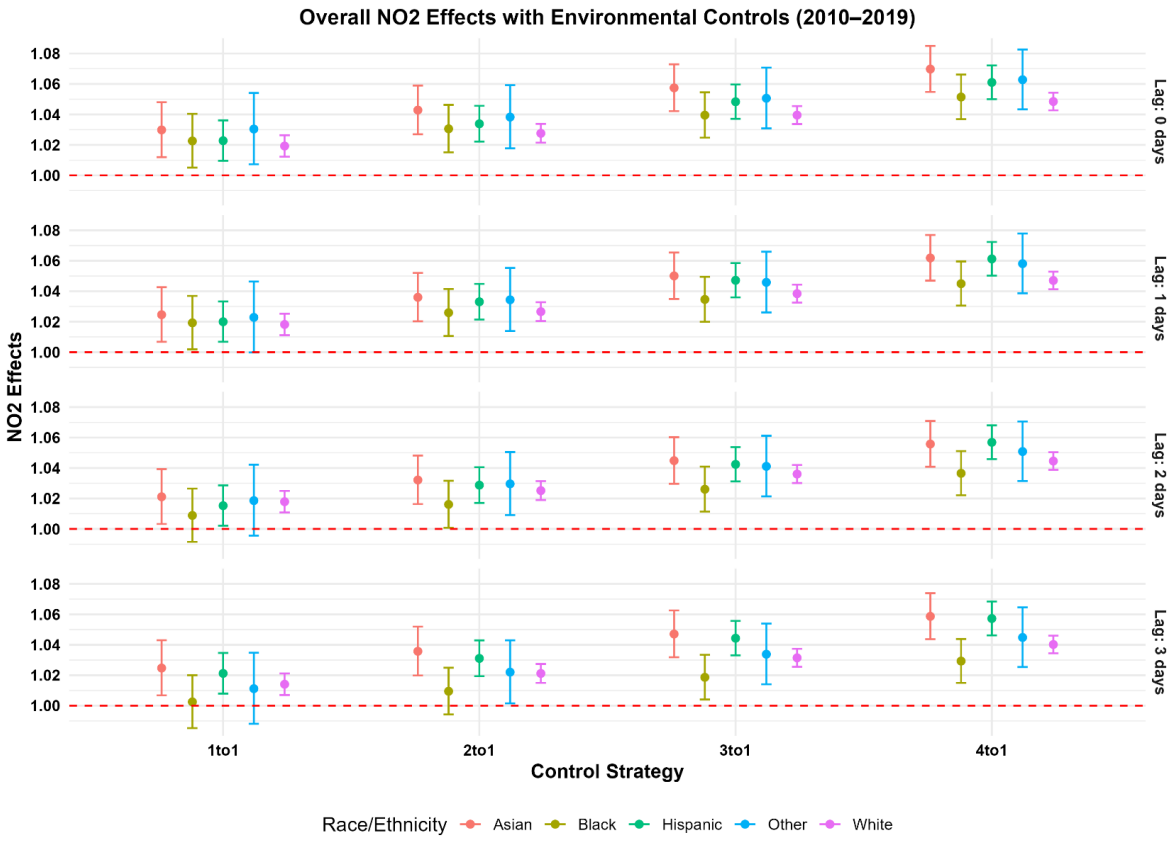


Figure 22. The overall impact of NO_2 exposure on hospital admissions stratified by race-ethnicity.

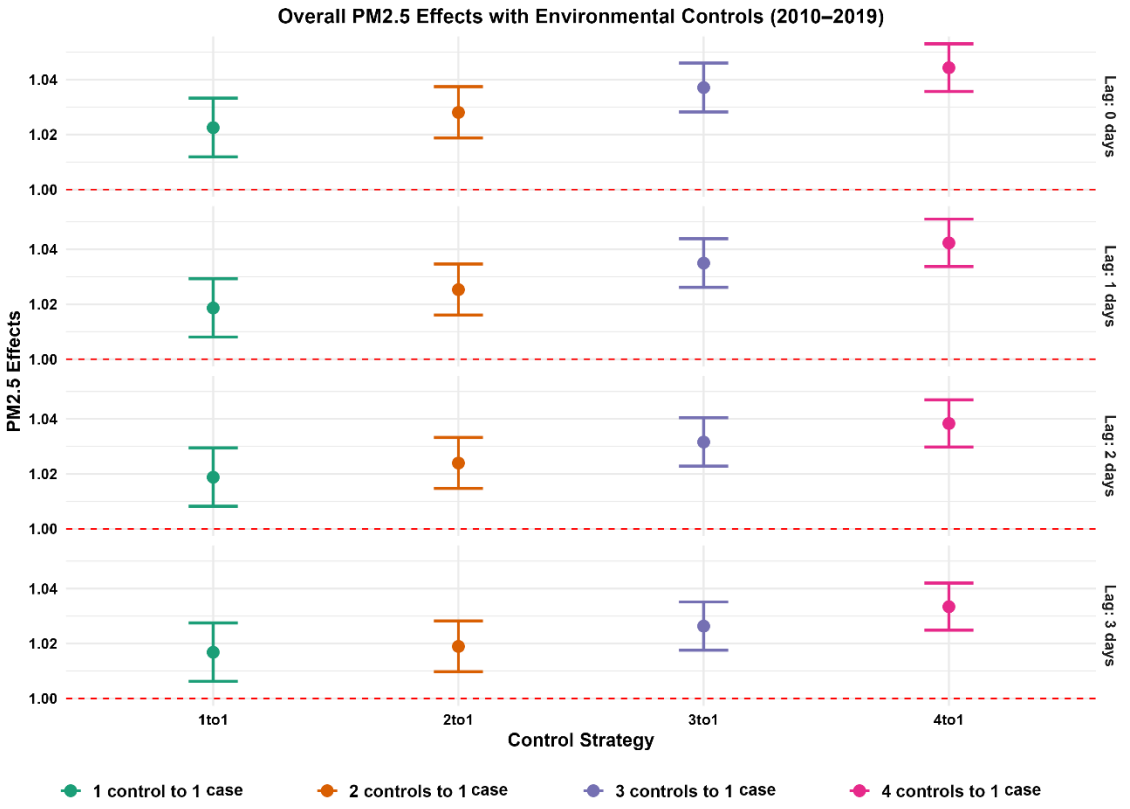


Figure 23. The overall impact of $PM_{2.5}$ exposure on hospital admissions.

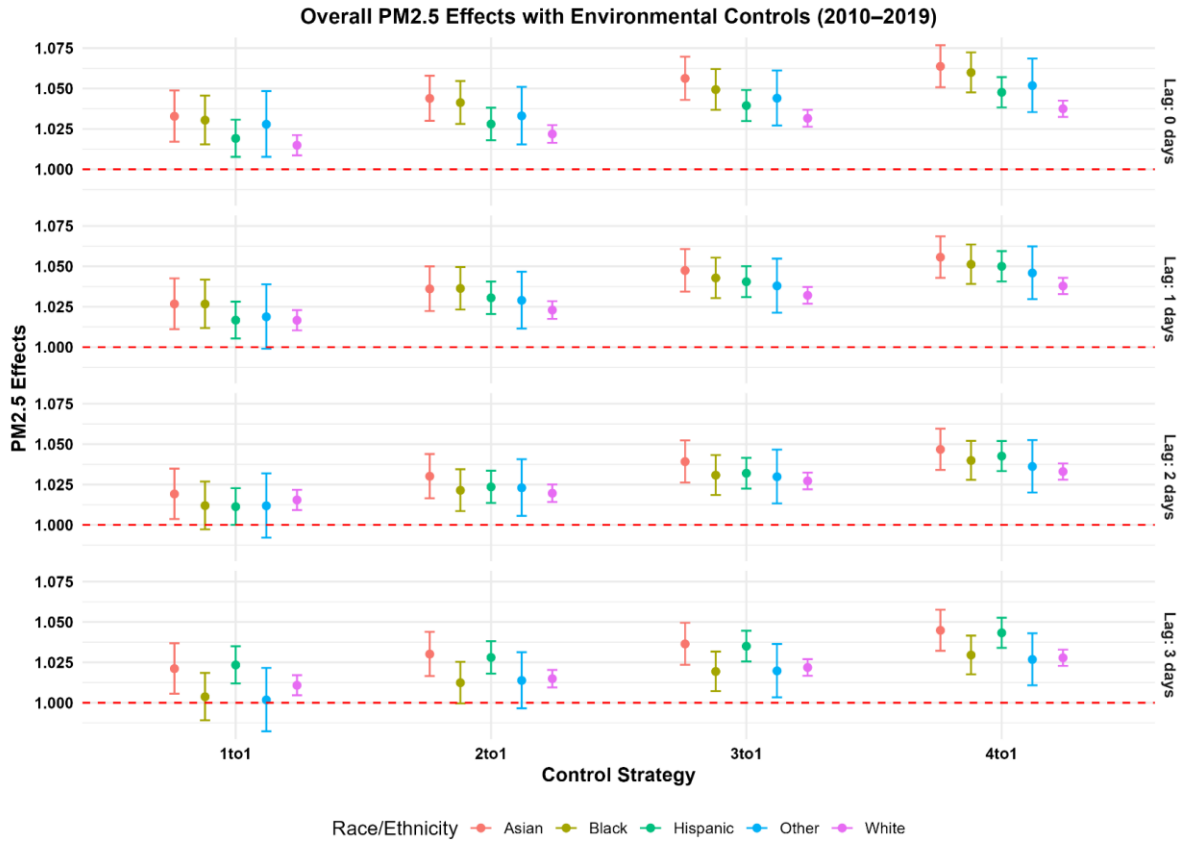


Figure 24. The overall impact of $PM_{2.5}$ exposure on hospital admissions stratified by race-ethnicity.

Modelling diabetes mellitus hospital LOS from NO_2 and $PM_{2.5}$ exposure using HCAI data

For impact of NO_2 on hospital LOS, exposure to higher daily concentrations of NO_2 was associated with a statistically significant increase in hospitalizations (Figure 25). The relationship was robust and consistent across all examined lag periods, confirming a clear and positive association between short-term NO_2 exposure and longer hospitalization durations. At lag 0, the mean ratio (MR) in LOS was 1.018 (95% CI: 1.017–1.019), representing the strongest effect across the lag structure. The association remained stable at lag 1 with an MR of 1.018 (95% CI: 1.017–1.019) and showed a gradual decline in magnitude at lag 2 (MR = 1.016, 95% CI: 1.015–1.017) and lag 3 (MR = 1.011, 95% CI: 1.010–1.012). Despite this attenuation, all associations remained statistically significant, indicating that elevated NO_2 levels continued to exert measurable influence on hospitalization duration for up to three days following exposure.

For impact of NO_2 on LOS by race and ethnicity, the stratified analyses demonstrated that the positive association between daily NO_2 exposure and increased hospitalizations was consistent across all major racial and ethnic groups, though the magnitude of the effect varied (Figure 26). Among Black patients, NO_2 exposure showed the strongest and most persistent relationship with LOS, with IRs ranging from 1.047 (95% CI: 1.037–1.058) at lag 0 to 1.033 (95% CI: 1.022–

1.043) at lag 3. These results indicate a robust and sustained increase in LOS up to three days after exposure. Similarly, White patients exhibited a consistent positive association, with IRs of 1.028 (95% CI: 1.024–1.033) at lag 0 and 1.019 (95% CI: 1.014–1.024) at lag 3, suggesting slightly smaller but still statistically significant effects. For Hispanic patients, the association strengthened with lag, peaking at lag 3 (MR = 1.042, 95% CI: 1.034–1.051), indicating that exposure-related impacts may extend over multiple subsequent days. In contrast, Asian patients demonstrated a more variable pattern, with a marked increase at lag 1 (MR = 1.052, 95% CI: 1.042–1.062) and lag 2 (MR = 1.035, 95% CI: 1.025–1.045), but a slight reduction below unity at lag 3, possibly reflecting population heterogeneity or smaller sample size. Finally, the Other race-ethnicity group showed highly stable and statistically significant effects across all lags, with IRs ranging from 1.021 to 1.022 for lags 0–2 and a moderate attenuation at lag 3 (MR = 1.016, 95% CI: 1.015–1.017).

For impact of $PM_{2.5}$ on LOS, analyses revealed a consistent and statistically significant association between short-term exposure and prolonged hospitalization (Figure 27). Across all lag periods examined (lag 0–3 days), elevated daily $PM_{2.5}$ concentrations were positively associated with longer hospital stays, indicating that short-term increases in ambient pollution levels contribute measurably to disease severity and extended inpatient recovery time. At lag 0, a one IQR increase in $PM_{2.5}$ was associated with a 0.84% longer LOS (MR = 1.008, 95% CI: 1.008–1.009), representing the strongest effect among the examined lags. The association modestly attenuated but persisted through subsequent days, showing statistically significant elevations at lag 1 (MR = 1.007, 95% CI: 1.006–1.008), lag 2 (MR = 1.005, 95% CI: 1.004–1.005), and lag 3 (MR = 1.006, 95% CI: 1.005–1.006).

For impact of $PM_{2.5}$ on LOS by race-ethnicity, the association remained statistically significant for most groups, though the magnitude and temporal pattern of effects varied across populations (Figure 28). Overall, the findings suggest that short-term exposure to fine particulate pollution can consistently prolong hospital recovery duration, with stronger effects observed among some racial and ethnic groups. For individuals classified as White, $PM_{2.5}$ exposure was consistently and significantly associated with longer hospital stays across all lag periods. The strongest association was observed at lag 0 (MR = 1.035, 95% CI: 1.031–1.039), indicating a 3.5% increase in LOS per IQR increase in $PM_{2.5}$. The effect slightly declined across lag 1–3 (MR range: 1.022–1.027) but remained statistically significant, suggesting a persistent short-term impact on hospitalization duration. Among Hispanic patients, $PM_{2.5}$ also demonstrated significant positive associations with LOS, with a clear increasing trend over time. The estimated effects rose from 1.016 (95% CI: 1.008–1.024) at lag 0 to 1.089 (95% CI: 1.081–1.097) at lag 3, indicating that prolonged exposure or delayed physiological responses may compound the effects of $PM_{2.5}$ in this population. This pronounced gradient highlights potential heightened vulnerability or delayed recovery among Hispanic patients. For Asian individuals, the associations were smaller: the effect at lag 0 (MR = 1.018, 95% CI: 1.009–1.028) indicated a

modest increase in LOS and the results at later lags were near or slightly below unity. In contrast, Black patients showed a distinct temporal pattern. At lags 0–1, the estimated MRs of expected LOS were slightly below 1 (e.g., lag 0 = 0.985, 95% CI: 0.976–0.994), indicating a modest reduction in expected length of stay immediately following exposure. In contrast, at lags 2–3, the associations became positive and statistically significant (e.g., lag 3 = 1.011, 95% CI: 1.003–1.020), corresponding to an increase in expected LOS. This may indicate a delayed manifestation of $PM_{2.5}$-related morbidity or differential care dynamics. Finally, among those categorized as Other, the association between $PM_{2.5}$ and LOS was both highly consistent and statistically significant across all lags, with nearly identical estimates around 1.010 (95% CI range: 1.009–1.011) from lag 0 to lag 3. This stability underscores a persistent, low-level increase in LOS attributable to particulate exposure across short-term windows.

Taken together, these findings confirm that short-term $PM_{2.5}$ exposure is significantly associated with extended hospital stays across all race-ethnicity groups, though the timing and magnitude of the effects vary, potentially reflecting differential exposure patterns, underlying health vulnerabilities, or access to care.

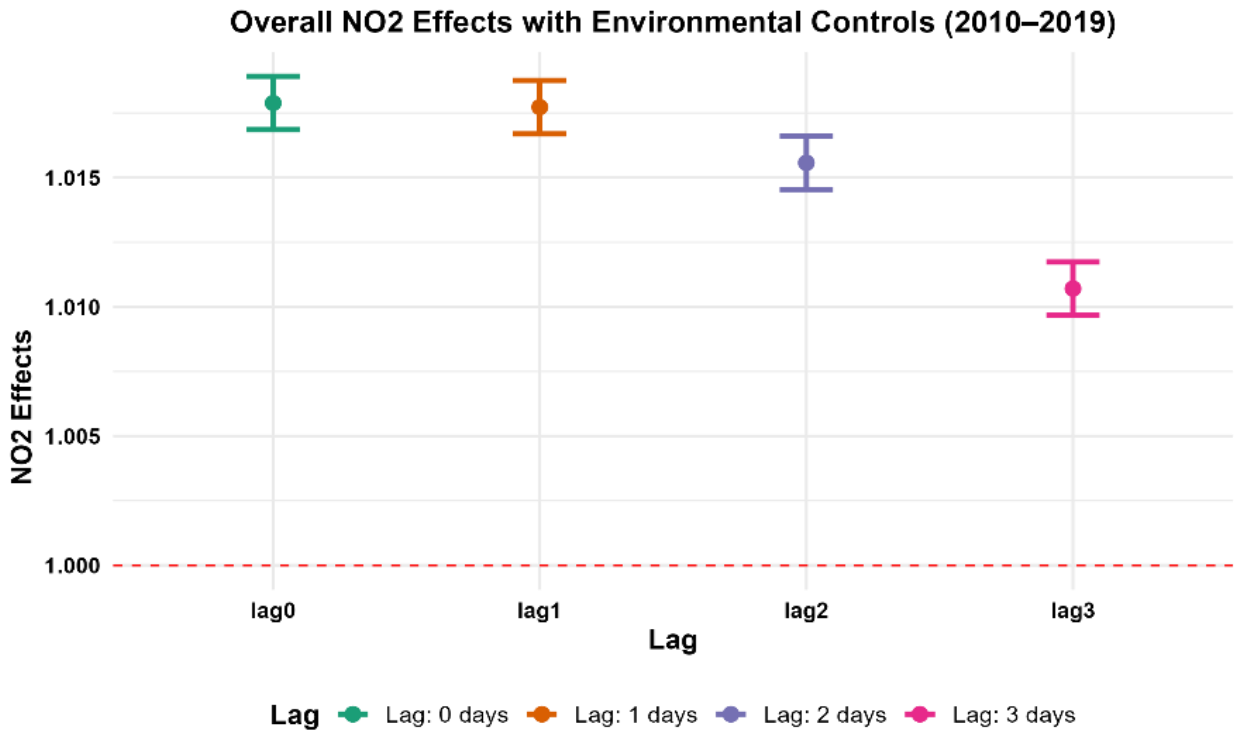


Figure 25. The overall impact of NO_2 exposure on hospital length of stay.

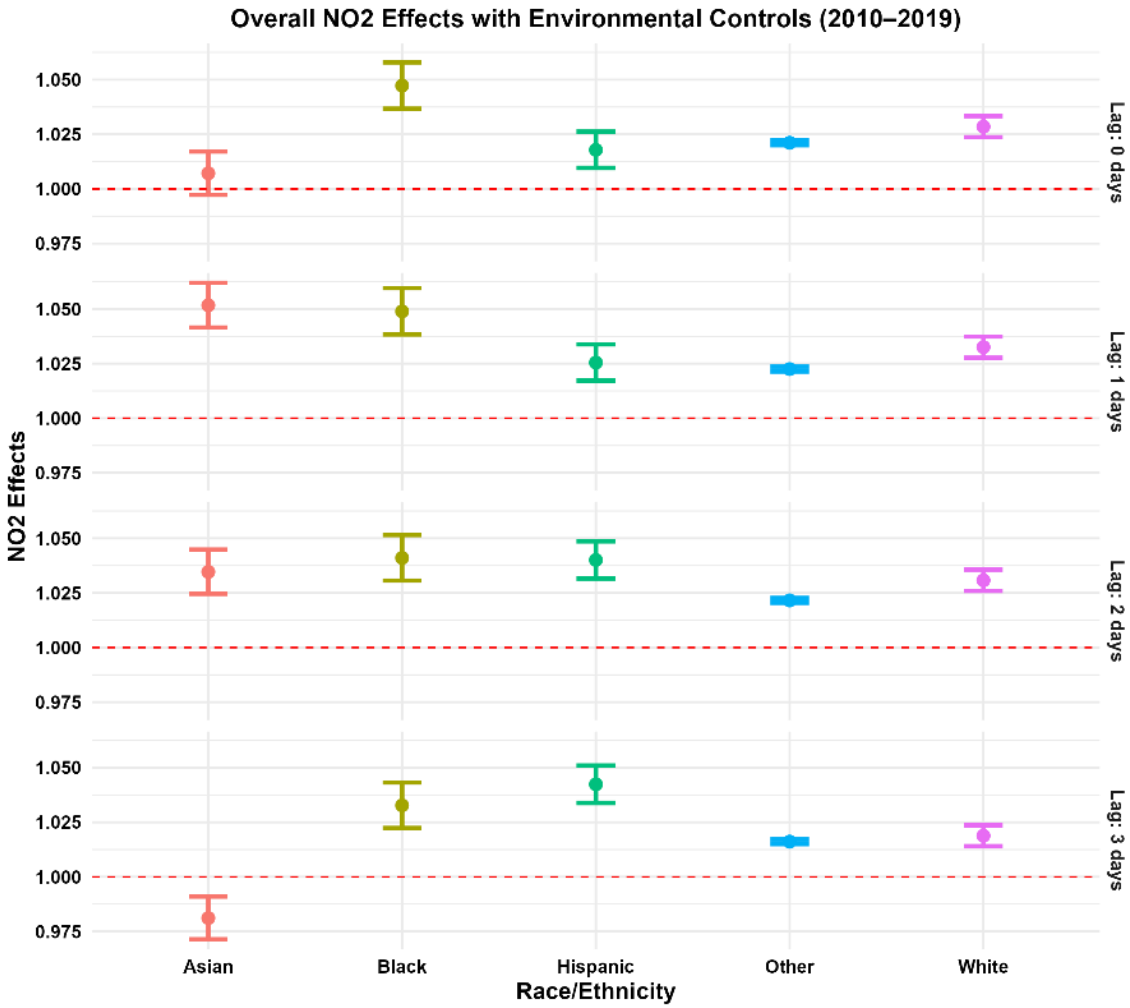


Figure 26. The overall impact of NO_2 exposure on hospital length of stay stratified by race-ethnicity.

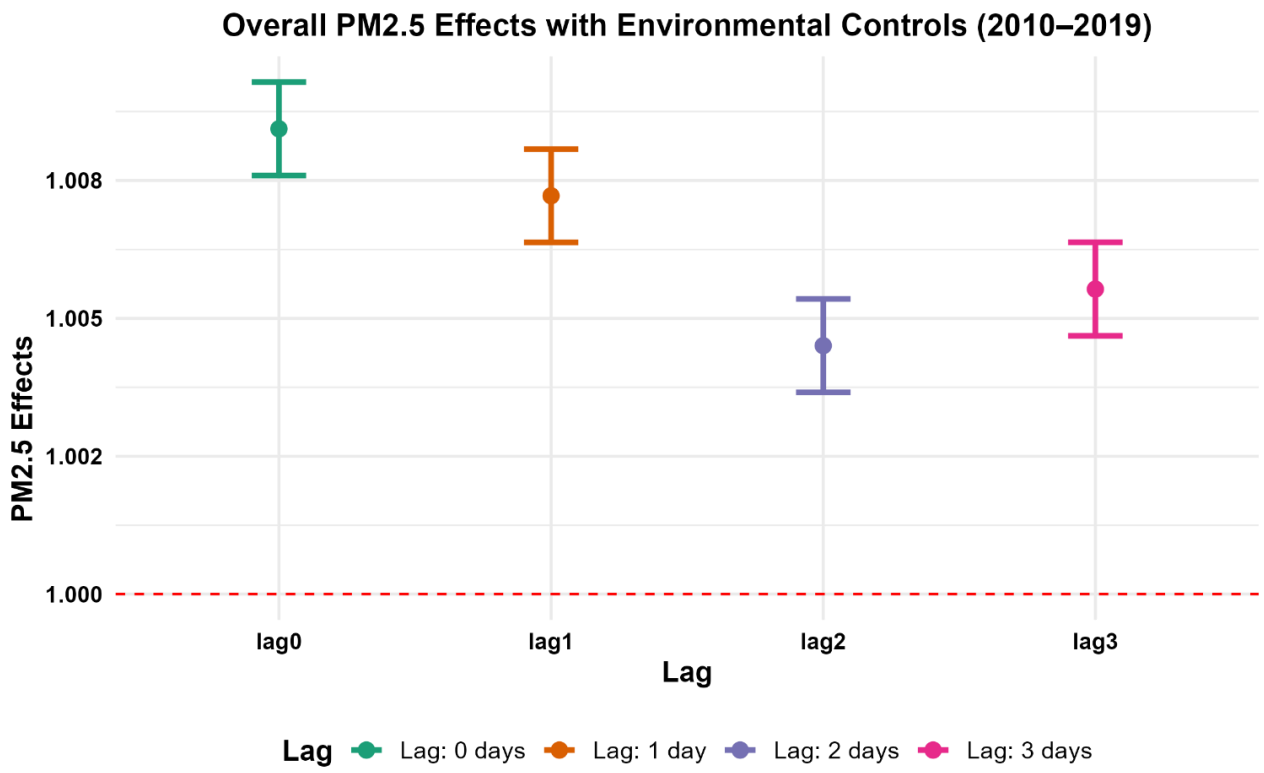


Figure 27. The overall impact of $PM_{2.5}$ exposure on hospital length of stay.

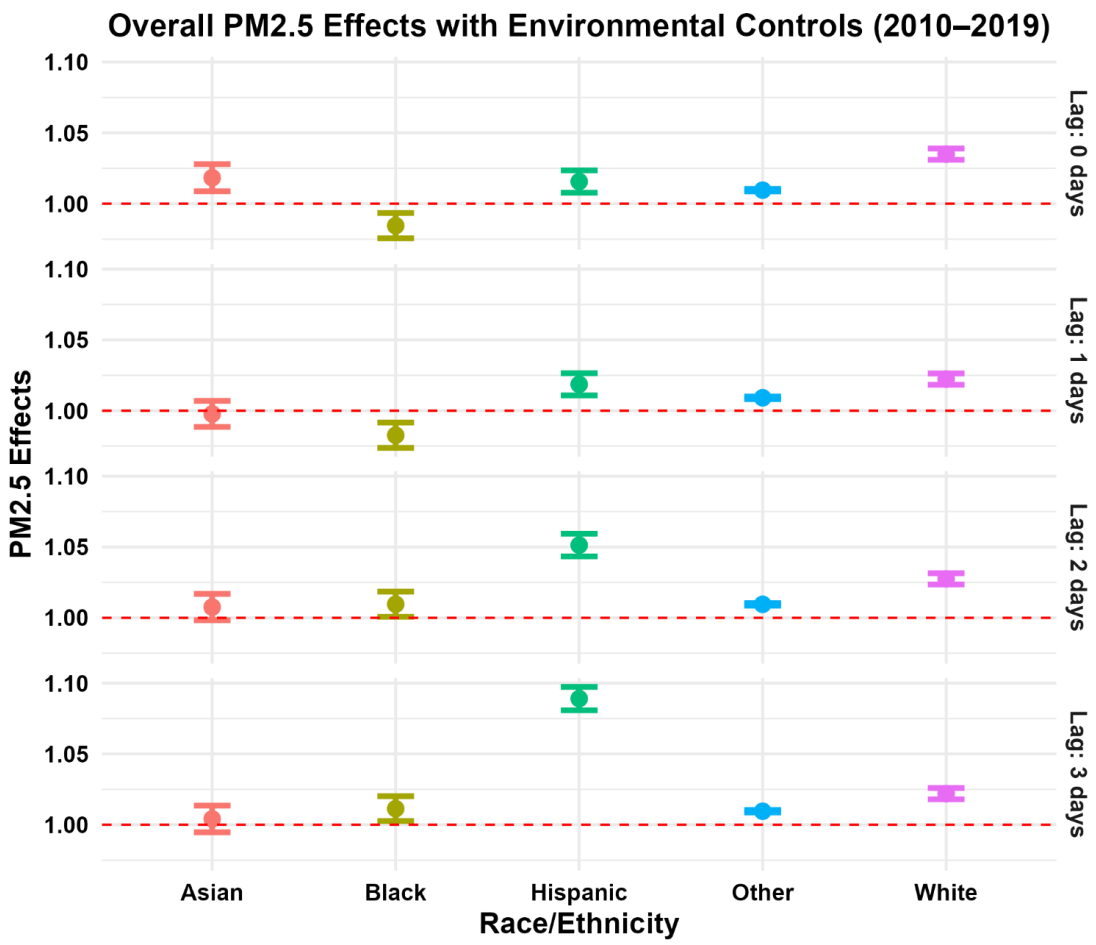


Figure 28. The overall impact of $PM_{2.5}$ exposure on hospital length of stay stratified by race-ethnicity.

Like ED visits, we also conducted inpatient analyses for individual years (2010-2019 for 10 years) across multiple stratification strategies, including health insurance pay type, region, primary language spoken, race-ethnicity, sex, and age group. Across all stratified analyses, the results consistently demonstrated a positive association between air pollution exposure, both NO_2 and $PM_{2.5}$, and adverse hospital outcomes, including increased admissions and longer length of stay. Detailed results of these stratified analyses are presented in Supplementary File 2.

Modelling diabetes mellitus ED visits and hospitalizations from ***O_3*** *exposure using HCAI data*

For T2D-related ED visits, O_3 exposure demonstrated small and generally inconsistent associations across lag periods and matching strategies event after controlling for impact from NO_2, $PM_{2.5}$ and socioeconomic status (Figure 29). At lag 0, the estimated odds ratios ranged narrowly from 0.994 to 1.002 across different control strategies, with most 95% confidence intervals overlapping unity, indicating little to no immediate effect. Similar patterns persisted through lag 1 to lag 3, with estimates fluctuating around the null (e.g., lag 1 range: 0.993–1.002;

lag 3 range: 0.995–1.000). A slight downward trend was observed under higher control ratios (3:1 and 4:1 matching), where ORs dipped modestly below 1.0, suggesting a potential inverse or null effect at those settings. Overall, the findings suggest that short-term variations in ambient O_3 were not consistently associated with increased ED utilization among individuals with T2D. This contrasts with the clearer and stronger associations seen for NO_2 and $PM_{2.5}$, indicating that O_3's impact on acute diabetic morbidity may be weaker, context-dependent, or confounded by co-pollutant interactions. The lack of a monotonic or lag-dependent pattern supports the interpretation that O_3 effects on acute healthcare utilization may be modest and transient in this population.

The estimated effects of O_3 on T2D-related hospital admissions were generally small and mixed across lags and control strategies (Figure 30). Effect estimates ranged approximately from 1.003 to 1.008, with most 95% confidence intervals spanning the null value. Significant associations were observed mainly under the more stringent control strategy (4-to-1 matching), where ORs reached 1.007 (95% CI: 1.001–1.013) at lag 1, 1.007 (95% CI: 1.001–1.013) at lag 2, and 1.008 (95% CI: 1.002–1.014) at lag 3, suggesting a modest but persistent positive relationship. In contrast, associations under less restrictive control strategies (1-to-1 or 2-to-1 matching) tended to be weaker and nonsignificant, reflecting greater heterogeneity and potential confounding. The modest increase in effect magnitude from lag 0 through lag 3 implies that O_3 may exert slightly delayed effects relative to other pollutants, consistent with prior studies reporting subacute respiratory or inflammatory responses that manifest over several days. Compared with NO_2 and $PM_{2.5}$, the overall O_3 effects were smaller and less consistent, likely due to the complex spatiotemporal behavior of O_3, its inverse correlation with traffic-related pollutants in urban cores, and differences in exposure misclassification across seasons. Nonetheless, the presence of significant associations at multiple lags under the stricter analytic design indicates that O_3 exposure may contribute incrementally to acute diabetes-related hospitalizations, albeit to a lesser extent than NO_2 and $PM_{2.5}$.

For hospital LOS, O_3 exposure showed a mixed temporal pattern with both positive and negative associations across lag days (Figure 31). At lag 0, O_3 was associated with a modest but statistically significant increase in LOS (MR = 1.018, 95% CI: 1.017–1.019), suggesting that same-day exposure may exacerbate disease severity or delay recovery among hospitalized T2D patients. However, this effect reversed at lag 1 (MR = 0.991, 95% CI: 0.990–0.992) and remained below unity at lag 2 (MR = 0.996, 95% CI: 0.995–0.996), indicating possible compensatory recovery or adaptive physiological responses following exposure. By lag 3, the association attenuated toward the null (MR = 0.999, 95% CI: 0.999–1.000). This oscillating pattern likely reflects the transient and complex biological response to O_3 exposure, where acute oxidative stress may initially worsen glycemic or inflammatory conditions, followed by short-term resolution or hospital treatment effects that mitigate impact on LOS. Similar bidirectional effects have been reported in previous studies, suggesting that O_3's influence on clinical

outcomes may depend on exposure timing, co-pollutant interactions, and individual susceptibility.

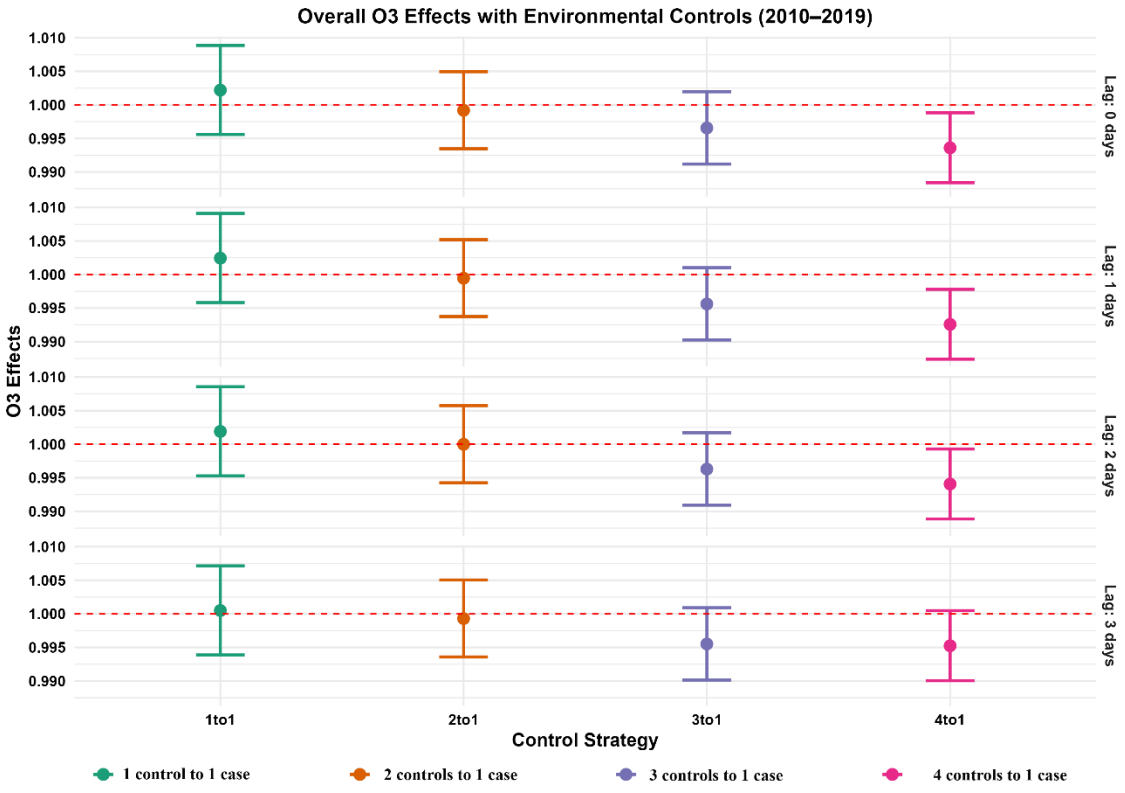


Figure 29. The impact of O_3 exposure on ED visits after adjusting for NO_2, $PM_{2.5}$ and socioeconomic status impacts.

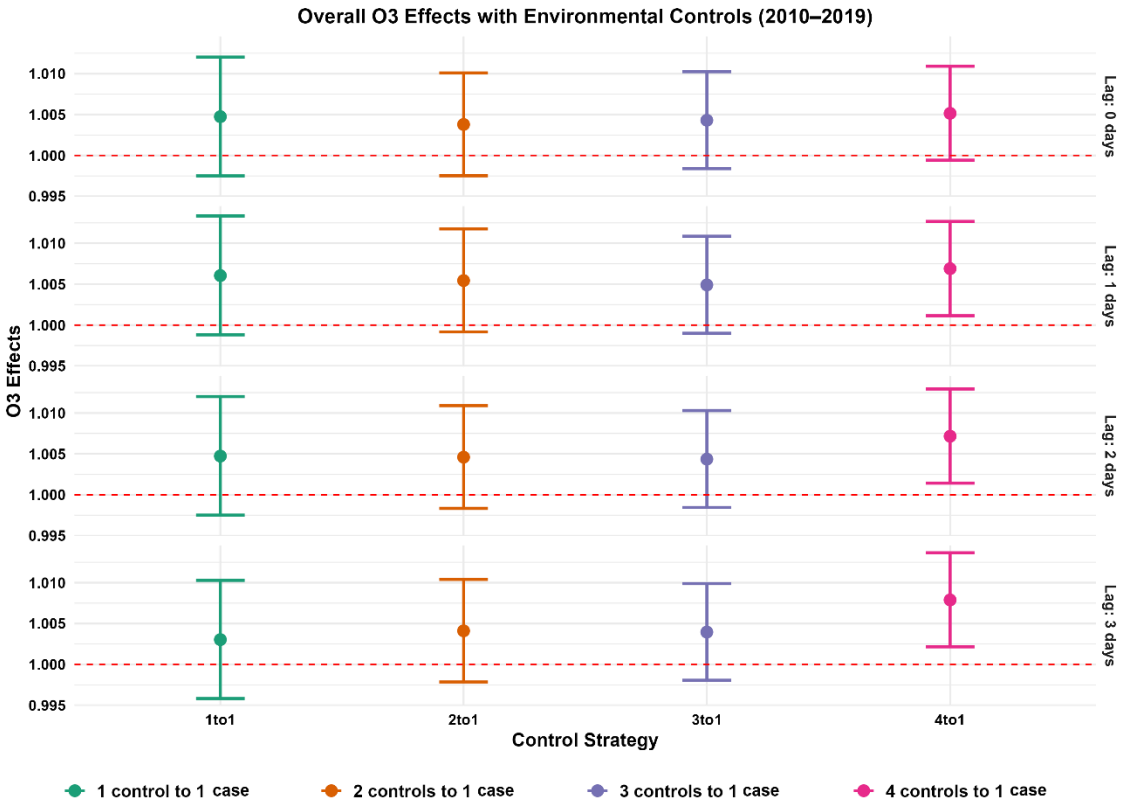


Figure 30. The impact of O_3 exposure on hospital admissions after adjusting for NO_2, $PM_{2.5}$ and socioeconomic impacts.

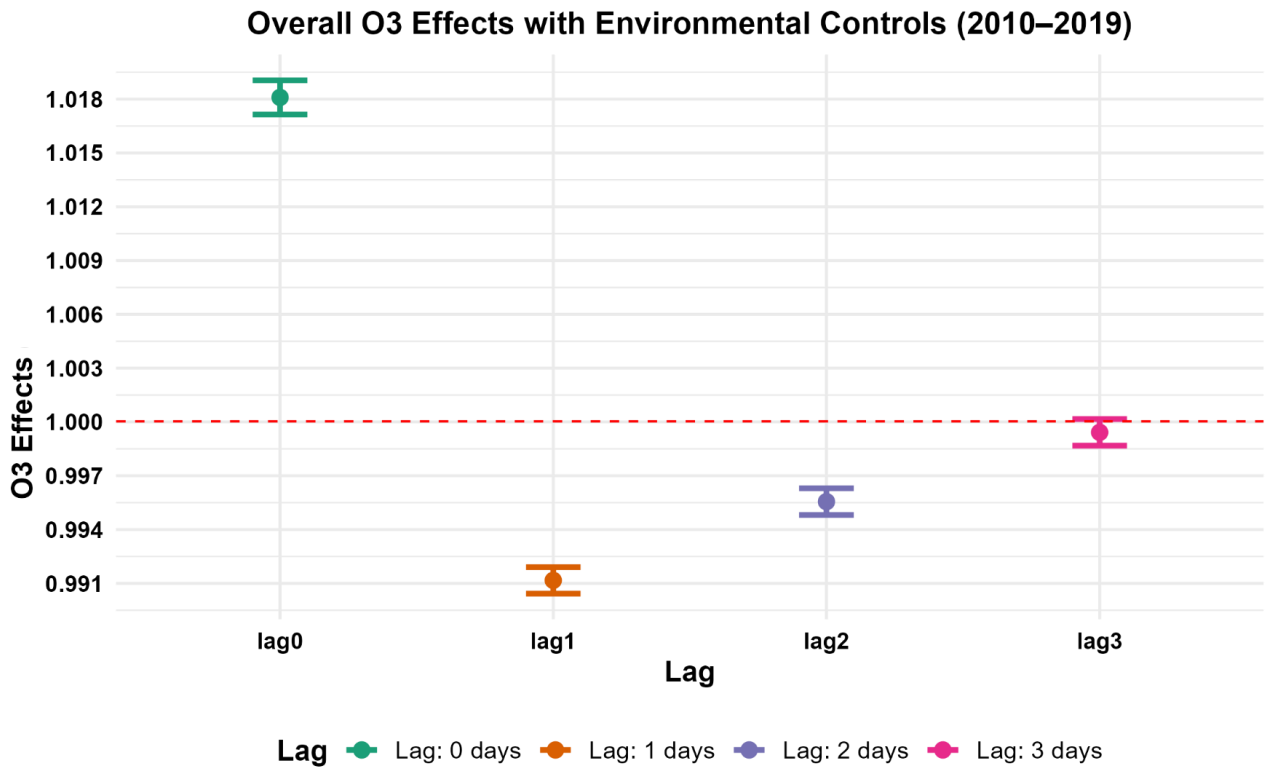


Figure 31. The impact of O_3 exposure on hospital LOS after adjusting for NO_2, $PM_{2.5}$ and socioeconomic status impacts.

Modeling diabetes mortality from NO_2 and $PM_{2.5}$ exposure using CDPH Vital Records data

Mortality analyses using CDPH data from 2010–2021 revealed statistically significant positive associations between exposure to ambient air pollutants and the odds of diabetes-related (primary cause: ICD10 E11) mortality among California residents. For NO_2, a 5.09 ppb IQR increase was associated with 3.6% higher odds of death (OR = 1.036; 95% CI: 1.010–1.063). The consistent positive association suggests that exposure to NO_2, an indicator of traffic-related pollution, contributes to increased mortality risk. In comparison, $PM_{2.5}$ showed a markedly stronger relationship: each 2.81 µg m^{-3} IQR increase in $PM_{2.5}$ exposure corresponded to 16.1% higher odds of death (OR = 1.161; 95% CI: 1.132–1.191). The larger magnitude and precision of the $PM_{2.5}$ effect highlights its dominant contribution to diabetes-related mortality, consistent with extensive epidemiologic evidence linking fine particulates to cardiovascular, respiratory, and metabolic dysfunction. Overall, both pollutants were associated with elevated diabetes-related mortality odds, but $PM_{2.5}$ exerted a substantially greater and more robust effect, emphasizing the need for continued air quality improvements targeting fine particulate sources to mitigate long-term health risks across California's population.

Task 6. Estimate economic benefits from reducing air pollution exposures on metabolic health outcomes

Based on the results of the preceding set of models for the pollutants NO_2, $PM_{2.5}$, and O_3, this section uses standardized medical expenditure and value of statistical life (VSL) data to determine the economic benefits of reducing T2D by reducing the pollutants NO_2, $PM_{2.5}$, and O_3 by an interquartile for hospitalization, emergency department and mortality outcomes (HCAI and CDPH data) or by ppb for NO_2 and O_3 or μg m^{-3} for $PM_{2.5}$ for medication outcome and T2D incidence outcomes (CHIS data). We present avoidable medical expenditures as well as the value of the overall avoidable loss of life.

Methodology

Constructing state-level economic burden requires integrating exposure–response functions (ERF) with cost-of-illness data. Using ERFs) presented in this report (each are indicated below when they are used), we determine potential impact fractions (PIF) from single pollutants to determine the incremental medical expenditures for T2D versus non-diabetic controls. Costs are adjusted to 2024 constant U.S. dollars to remove the effects of inflation.

Potential impact fractions (PIFs) are the proportion of cases that are likely to be prevented if the exposure to the pollutant in question were eliminated by a particular proportion, assuming the observed association is unbiased.[129] To do this, we assume the odds-ratios estimated for logistic case-control and conditional logistic case-control models are unbiased (e.g., each logistic or conditional logistic model is sufficiently adjusted) and that the outcome is sufficiently rare (e.g., $\cong 10\%$) such that the relevant odds-ratio does not significantly overstate the risk-ratio such that $RR_{IQR} \approx OR_{IQR}$, where OR_{IQR} is the odds-ratio and RR_{IQR} is the risk-ratio for an interquartile change in a given pollutant.[130-132] The formula to calculate this can be derived from the equation of a distribution shift PIF where the distribution shift is uniform (e.g., interquartile), assuming $RR_{IQR} \approx OR_{IQR}$: [129]

$$PIF = 1 - \frac{1}{OR_{IQR}}$$

where OR_{IQR} is the odds-ratio for an interquartile change in a given pollutant as estimated from equations in the previous sections above. For results using CHIS data, the odds-ratio only represents a change in ppb (for NO_2 and O_3) or μg m^{-3} for $PM_{2.5}$) See Supplementary File 5 for the equation derivation. To determine economic benefits, we then multiply the relevant PIF by the relevant medical expenditure or value of a statistical life.[133-139] This can be expressed as

$$AME = PIF\ x\ \widehat{ME}$$

where *AME* is avoidable medical expenditure, and $\widehat{ME}$ is mean medical expenditure in California. We compute standard errors for *AME* based on the principles of the delta method as shown in Supplementary File 6.

Data

Medical Expenditures

Medical expenditure estimates data are based on recently released expenditure data for the years 2010-2019.[140] All data are available from the Institute for Health Metrics and Evaluation,[142] and the code used to develop the estimates is available via GitHub (https://github.com/ihmeuw/Resource_Tracking_US_DEX/tree/main/DEX_Capstone_2025). To compile this dataset, multiple administrative data sources were used that together reflect approximately 40 billion insurance claims and approximately 1 billion facility records across Medicare (including Medicare Advantage), Medicaid, private insurance, and out-of-pocket payments.[140] Data were drawn from seven major medical claims sources including MarketScan, Kythera, the Health Care Cost Institute, the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project, and the Medical Expenditure Panel Survey.

Expenditure data were classified by patient age, sex, type of care, payer, and county of residence. Type of care was classified as ambulatory, emergency department, home health, inpatient, nursing facilities, and medications. Payers were classified as Medicare, Medicaid, private, and out-of-pocket. Diagnoses were mapped to 148 standardized health conditions using ICD-9/10 codes and National Drug Codes, following the Global Burden of Disease 2019 framework including T2D.[140] Statistical adjustments were applied to correct for incomplete data (e.g., facilities reporting only charges) and to reallocate spending to comorbidities using penalized linear regression models.

We only used California-specific age-standardized expenditures for our analysis, adjusted to 2024 constant US dollars. The ERFs used in this analysis come from equations that each produced a single OR_{IQR} per equation, so we use average annual medical expenditures in our analysis to determine the average annual potential medical expenditure savings due to reductions in each pollutant. We also present spending per capita, encounters per capita, and spending per encounter. The relevant estimates used are presented in Tables 15-18.

Table 15. Total Spending Summary Statistics for State-Level Medical Expenditure Data ($2024)

	All Spending ($billion)			Emergency Department ($million)			Inpatient ($billion)			Prescribed Pharmaceuticals ($billion)		
Year	Mean	95%	CI	Mean	95%	CI	Mean	95%	CI	Mean	95%	CI
2010	17	16	17	236	223	236	1.4	1.2	1.5	4.1	3.8	4.2
2011	19	17	19	248	223	273	1.5	1.4	1.7	4.1	3.8	4.3
2012	19	17	20	260	236	285	1.5	1.4	1.7	4.1	4.0	4.3
2013	19	17	20	273	248	298	1.5	1.4	1.7	4.2	4.0	4.5
2014	20	19	20	273	248	298	1.5	1.4	1.6	4.7	4.6	5.0
2015	21	20	21	298	285	310	1.7	1.6	1.7	5.0	4.8	5.2
2016	21	21	22	310	298	322	1.7	1.6	1.9	5.0	4.8	5.2
2017	21	20	22	310	298	347	1.7	1.5	2.0	5.0	4.6	5.1
2018	21	20	22	322	285	347	1.9	1.6	2.1	5.0	4.7	5.2
2019	21	21	22	360	335	372	2.0	2.0	2.1	5.1	5.0	5.3

Table 16. Spending Per Capita Summary Statistics for State-Level Medical Expenditure Data ($2024)

	All Spending ($)			Emergency Department ($)			Inpatient ($)			Prescribed Pharmaceuticals ($)		
Year	Mean	95%	CI	Mean	95%	CI	Mean	95%	CI	Mean	95%	CI
2010	581.32	557.95	604.91	7.80	7.46	8.08	45.87	42.42	49.79	136.69	129.85	143.65
2011	603.98	571.78	636.67	8.30	7.58	8.89	50.47	45.52	56.59	135.42	127.14	143.65
2012	600.16	569.32	638.22	8.48	7.82	9.50	49.87	43.88	55.89	135.22	129.07	142.40
2013	603.02	575.40	638.20	8.70	8.00	9.63	49.41	45.02	56.10	137.85	130.36	145.86
2014	630.40	613.35	645.26	8.74	8.06	9.55	48.82	45.52	52.94	153.30	146.61	161.26
2015	667.37	649.60	682.15	9.42	9.01	9.93	54.35	51.60	57.18	159.60	153.16	165.66
2016	680.95	664.09	695.57	9.88	9.46	10.43	54.67	51.92	57.51	158.48	152.52	164.26
2017	648.27	614.40	694.02	9.96	9.30	10.79	53.68	46.19	62.22	154.42	145.84	160.43
2018	648.84	608.52	705.32	9.91	8.98	10.80	57.04	49.13	67.44	155.46	147.15	164.66
2019	671.41	650.28	691.45	11.01	10.50	11.61	62.47	59.81	65.00	160.43	152.82	164.96

Table 17. Encounters Per 1,000 Population Summary Statistics for State-Level Medical Expenditure Data

	Emergency Department			Inpatient			Prescribed Pharmaceuticals		
Year	Mean	95%	CI	Mean	95%	CI	Mean	95%	CI
2010	6.80	6.07	7.63	2.15	1.99	2.38	821.43	768.54	872.06
2011	6.48	5.06	8.26	2.19	1.90	2.63	817.38	732.18	914.77
2012	6.28	4.99	7.99	2.11	1.83	2.61	785.44	695.82	891.73
2013	6.02	4.90	7.69	2.05	1.76	2.39	758.78	673.19	840.66
2014	5.93	5.43	6.40	1.93	1.81	2.11	792.14	729.85	864.79
2015	6.83	6.46	7.20	2.17	2.11	2.26	819.90	776.12	869.80
2016	7.09	6.72	7.48	2.14	2.09	2.23	824.83	783.30	876.67
2017	7.02	5.64	8.76	2.04	1.76	2.46	742.30	642.47	869.41
2018	6.63	5.00	8.43	2.09	1.81	2.45	691.36	602.22	792.48
2019	7.28	6.73	7.65	2.19	2.07	2.29	675.07	635.76	703.76

Table 18. Spending Per Encounter Summary Statistics for State-Level Medical Expenditure Data ($2024)

	Emergency Department ($)			Inpatient ($)			Prescribed Pharmaceuticals ($)		
Year	Mean	95%	CI	Mean	95%	CI	Mean	95%	CI
2010	770.60	664.27	857.70	18438.30	17313.38	19423.86	173.45	165.42	185.02
2011	894.67	690.38	1139.96	20110.94	17692.69	22270.77	173.41	163.05	183.40
2012	980.82	682.21	1286.62	20584.50	17074.92	23109.88	180.37	168.48	190.96
2013	1056.95	822.03	1367.37	21001.26	18492.12	23498.99	190.18	178.19	199.79
2014	1066.51	953.73	1193.76	21808.38	20681.71	23596.21	204.34	195.15	218.93
2015	988.65	916.94	1061.23	21113.23	19874.10	22491.00	212.51	199.37	226.60
2016	1006.27	933.41	1078.69	21524.42	20283.80	22955.00	213.01	199.70	227.53
2017	1100.34	849.60	1457.97	22848.74	20243.12	25775.51	226.29	212.85	240.99
2018	1133.51	883.55	1599.04	23621.26	21483.00	25983.33	228.41	215.86	248.94
2019	1055.15	988.19	1156.83	24259.11	22843.28	25612.94	234.06	221.35	245.98

Value of a Statistical Life

We use a standard Value of a Statistical Life (VSL), adjusted to 2024 constant US dollars, from CARB.(https://ww2.arb.ca.gov/sites/default/files/2021-10/SCAQMD%20Mortality%20Risk%20Reduction%20Valuation.pdf). This value is $14.5 million.

Mortality due to T2D

We used CDC Wonder to determine the average number of deaths in California due to T2D from 2010 to 2019 (https://wonder.cdc.gov/ucd-icd10.html). We included ICD-10 Diagnosis Codes E11-E14, excluding E10 (code for insulin-dependent diabetes). The age-adjusted average rate for California across 2010-2019 was 20.07 per 100,000 per year.

Results

Table 19 presents the results of changes in incidence of T2D based on the previously presented models based on data from the California Health Interview Survey (CHIS). These analyses examine overall incidence since the estimated OR_{IQR} does not differentiate by insurance type. To determine the medical expenditure impact, total costs for all categories of care (ambulatory, emergency department, home health, inpatient, nursing facilities, and medications) related to T2D were aggregated and used to determine the average annual medical expenditures for T2D attributed to each pollutant. Overall data can be expressed as total spending, spending per capita, encounters per capita, and spending per encounter.

Table 19. Avoidable Medical Spending/Use from Reduced Pollution, by Pollutant

Pollutant	PIF	Estimated Savings from a Unit Reduction of Pollutant (2024$)	
		Total Spending ($millions) (95% CI)	Spending Per Capita ($) (95% CI)
NO_2	0.012	$245.29 (150.75, 339.82)	$7.83 (4.81, 10.85)
$PM_{2.5}$	0.072	$1421.11 (1186.94, 1655.28)	$45.38 (37.90, 52.86)
O_3	0.060	$1193 (737.88, 1650.01)	$38.13 (23.56, 52.69)

Note: OR estimates used here represent changes in units (ppb for NO_2 and O_3, μg/m^3 for $PM_{2.5}$) (NO_2 1.0125, 95% CI:1.0073, 1.0170; $PM_{2.5}$ 1.0771, 95% CI: 1.0640, 1.0904; O_3 1.064, 95% CI: 1.090, 1.039) come from the highest reported OR for each pollutant listed in the CHIS Incidence Criteria Pollutants section of Supplementary File 4.

The next set of results, Table 20-Table 22 present impacts of pollutants on various subsets of medical expenditures, including medication, emergency department use, and inpatient care. These are used since we have separately modeled estimates for these subcategories of medical care utilization. Medication models were based on CHIS data, and emergency department and inpatient use models were based on data from HCAI. Finally, Table 23 presents the average annual VSL of avoidable deaths per due to T2D.

Table 20. Avoidable Emergency Department Spending/Use from Reduced Pollution, by Pollutant

Pollutant	PIF	Estimated Savings from Interquartile Reduction of Pollutant (2024$)			
		Total Spending ($millions) (95% CI)	Spending Per Capita ($) (95% CI)	Encounters Per Capita (95% CI)	Spending Per Encounter ($) (95% CI)
NO_2	0.007	$2.07 (1.17, 2.97)	$0.07 (0.04, 0.09)	0.05 (0.03,0.07)	$7.22 (3.84, 10.59)
$PM_{2.5}$	0.006	$1.83 (1.08, 2.58)	$0.06 (0.03, 0.08)	0.04 (0.03, 0.06)	$6.37 (3.53, 9.22)
O_3	0.002	$0.70 (-1.21, 2.61)	$0.02 (-0.04, 0.08)	0.02 (-0.03, 0.06)	$2.44 (-4.21, 9.10)

Note: OR_{IQR}estimates used here (NO_2 1.0072, 95% CI:1.0041, 1.0103, IQR: 7.64 ppb, mean: 9.93 ppb; $PM_{2.5}$ 1.0063, 95% CI: 1.0037, 1.0089, IQR: 3.65 μg/m³, mean: 9.59 μg/m³; O_3 1.0024, 95% CI: 0.9958, 1.0090; IQR: 10.56 ppb, mean: 35.58 ppb) come from the highest reported OR_{IQR} for each pollutant listed in the HCAI ED Overall section of Supplementary File 4.

Table 21. Avoidable Inpatient Spending/Use from Reduced Pollution, by Pollutant

Pollutant	PIF	Estimated Savings from Interquartile Reduction of Pollutant (2024$)			
		Total Spending ($millions) (95% CI)	Spending Per Capita ($) (95% CI)	Encounters Per Capita (95% CI)	Spending Per Encounter ($) (95% CI)
NO_2	0.052	$86.11 (69.24, 102.98)	$2.77 (2.23, 3.31)	0.11 (0.08, 0.13)	$1132 (913, 1352)
$PM_{2.5}$	0.042	$69.46 (54.87, 84.05)	$2.23 (1.77, 2.70)	0.08 (0.07, 0.10)	$914 (724, 1104)
O_3	0.008	$12.81 (3.46, 22.16)	$0.41 (011, 0.71)	0.02 (0.004, 0.03)	$169 (46, 291)

Note: OR_{IQR} estimates used here (NO_2 1.0555, 95% CI:1.0455, 1.0656, IQR: 7.70 ppb, mean: 8.53ppb; $PM_{2.5}$ 1.0443, 95% CI: 1.0356, 1.0530, IQR: 3.57 µg/m³, mean: 8.64 µg/m³ ; O_3 1.0078, 95% CI: 1.0021, 1.0136, IQR: 10.29 ppb; mean: 38.43 ppb) come from the highest reported OR_{IQR} for each pollutant listed in the HCAI IP Overall section of Supplementary File 4.

Table 22. Avoidable Medication Spending/Use from Reduced Pollution, by Pollutant

Pollutant	PIF	Estimated Savings from a Unit Reduction of Pollutant (2024$)			
		Total Spending ($millions) (95% CI)	Spending Per Capita ($) (95% CI)	Encounters Per Capita (95% CI)	Spending Per Encounter ($) (95% CI)
NO_2	0.003	$85.04 (61.67, 108.40)	$2.74 (1.99, 3.50)	14.25 (10.16, 18.34)	$3.75 (2.71, 4.80)
$PM_{2.5}$	0.005	$315.15 (270.99, 359.32)	$10.16 (7.35, 12.96)	52.81 (44.23, 61.40)	$13.91 (11.88, 15.94)
O_3	0.034	$15.82 (12.46, 19.19)	$5.10 (2.97, 7.23)	26.51 (20.46, 32.56)	$6.98 (5.47, 8.50)

Note: OR estimates used here represent changes in units (ppb for NO_2 and O_3, µg/m³ for $PM_{2.5}$) (NO_2 1.0187, 95% CI:1.0136, 1.0239; $PM_{2.5}$ 1.0733, 95% CI: 1.0629, 1.0838; O_3 1.0355, 95% CI: 1.0279, 1.0431) come from the highest reported OR for each pollutant listed in the CHIS Medication Criteria Pollutant section of Supplementary File 4.

Table 23. Average Annual Value of Statistical Lives (VSL) Lost Due to T2D, due to an Interquartile Change in Pollution, by Pollutant

Pollutant	Avoidable Lost VSL per 100,000 population from Reducing Pollutant by an Interquartile ($millions)
NO_2	$10.379 (3.006, 17.753)
$PM_{2.5}$	$41.422 (34.886, 47.958)

Note: OR_{IQR} estimates used here (NO_2 1.036, 95% CI:1.010, 1.063, IQR: 5.09 ppb; $PM_{2.5}$ 1.161, 95% CI: 1.132, 1.191, IQR: 2.81 µg/m³) come from the section above *Modeling diabetes mortality from NO_2 and $PM_{2.5}$ exposure using CDPH Vital Records data*.

Limitations

Note that the pollution levels in each interquartile are not consistent across time but rather reflect relative movements within each distribution of pollutant levels, where the distribution is changing over time. Thus, an interquartile movement within a narrower distribution of pollutant levels is less than an interquartile movement within a wider distribution of pollutant levels. The results are thus analyzing the average interquartile movement across the entire period, not specific levels of pollutants across the entire period. The models also model one pollutant at a time and assume no confounding from omitting other pollutants within each model.

Regarding other cost information, including the cost of caregiving, productivity losses (e.g., loss of employment, works days lost due to illness or medical treatment, etc.), and changes in subjective health status due to T2D, estimates of the impact of pollutants on each of these, via changes in the incidence of T2D, would be needed, but were not available.

No information could be provided on a per beneficiary basis. Beneficiaries are defined based on common insurance type and no equations were estimated by insurance type.

Regarding the assumption of unbiasedness that underlies the calculation of PIFs, a PIF is a counterfactual quantity (the proportion of cases that would be prevented under a specified intervention on exposure) so its identification requires strong assumptions beyond statistical significance, including consistency, positivity, and exchangeability/no unmeasured confounding in the underlying population. In a matched case–control framework estimated via conditional logistic regression (CLR), the exposure coefficient is identified from within–matched-set contrasts; matching does not itself eliminate confounding and can induce selection/collider bias, which is why conditioning on the matched structure and adjusting for measured confounders is necessary.[143-145] While we adjust for a broad set of observed individual and contextual covariates, residual confounding by unmeasured or mismeasured time-varying factors (e.g., comorbidities, health-care utilization shocks, co-exposures) may remain and could bias both the odds ratios and the implied PIFs. Moreover, CLR estimates a conditional odds ratio, whereas PIF is fundamentally risk-based and population-level. Since odds ratios are non-collapsible and effects may be heterogeneous across strata, conditional and marginal effects can differ even without confounding, so PIFs can be sensitive to how conditional estimates are mapped/standardized to the target population.[146-148] Finally, PIF magnitudes can be affected by model specification (functional form and interactions), exposure definitions, and finite-sample/sparse-data bias in logistic-type estimators,[149,150] and PIF calculations from case–control data are sensitive to these assumptions.[151-154] Nevertheless, our statistical approach has sought to minimize these sources of bias.

Discussion

In this project, diabetes-related health outcome modeling was implemented through a staged and prioritized analytic framework. We first focused on establishing robust concentration-

response relationships for single-pollutant exposures across five metabolic health endpoints, including diabetes incidence, medication use, diabetes-related emergency department visits, hospitalizations, length of stay, and mortality. These analyses were conducted not only for the full study period but also through year-specific models to characterize temporal heterogeneity. The scope and depth of these analyses, especially the combination of individual-level data, lagged exposure structures, and extensive stratification, required substantially greater effort than originally anticipated and ultimately exceeded the analytical capacity supported by the awarded budget. Using the CHIS, HCAI, and CDPH data collectively allowed a comprehensive assessment of both annual long-term and daily acute effects, capturing distinct but complementary health outcomes.

Exposure modeling in this project provides the critical exposure foundation that underpins all subsequent epidemiologic and policy-relevant analyses in this project. Task 2 developed high-resolution daily air pollution surfaces for criteria pollutants using an integrated machine learning–enhanced land use regression framework that combined regulatory monitoring, mobile Google Street View observations, and a rich set of spatial-temporal predictors. These models demonstrated strong explanatory performance and enabled assignment of daily exposures at fine spatial scales across California. The resulting exposure surfaces substantially reduced exposure misclassification relative to reliance on fixed-site monitoring alone and allowed for robust linkage with individual-level health records in Task 5, particularly for short-term outcomes such as emergency department visits, hospitalizations, and length of stay.

The relatively lower adjusted R_2 for the $PM_{2.5}$ model (0.65) still corresponds to a strong overall correlation (exceeding 0.80) between predicted and observed concentrations, indicating strong model performance. Compared with NO_2, $PM_{2.5}$ is influenced by a broader set of sources and processes, including regional transport, secondary formation, and episodic events such as wildfires, which are less tightly coupled to local land-use and traffic predictors typically used in LUR frameworks. Although additional regional or synoptic-scale predictors could further increase explanatory power, our primary objective was to characterize small-area spatial variability rather than maximize total variance explained. Accordingly, the model prioritizes local-scale predictors that are most relevant for exposure assignment in epidemiologic analyses.

In our exposure modeling, the large numerical magnitude of the season coefficients reflects the scale and coding of the seasonal indicator variables rather than an outsized physical effect. Season was modeled as a binary indicator (0/1), so its contribution to predicted concentrations is either zero or approximately the coefficient value (≈356–360). In contrast, continuous predictors such as temperature have much smaller coefficients but operate over a substantially larger range (e.g., >300 Kelvin), yielding comparable or larger contributions to the linear predictor when multiplied by their observed values. Seasonal indicators in LUR models implicitly capture multiple unmeasured or partially measured processes that vary systematically over the year,

including seasonal emission patterns (e.g., heating-related combustion, traffic activity), atmospheric chemistry, boundary-layer dynamics, and photochemical conditions that are not fully represented by individual meteorological variables. We explicitly evaluated multicollinearity during the D/S/A modeling process and retained seasonal terms only when they remained statistically significant and improved predictive performance after accounting for meteorology and other predictors. The D/S/A model controlled for collinearity, suggesting that the seasonal terms capture residual temporal structure rather than inflating coefficients through redundancy.

We extended this exposure modeling framework from criteria pollutants to air toxics, addressing a major gap in statewide exposure assessment. By applying D/S/A LUR techniques and developing annual concentration surfaces for multiple hazardous air pollutants, this task characterized long-term spatial variability in air toxics exposures across California communities. Although the annual temporal resolution limited the application of air toxics exposures to acute outcomes, these models were essential for evaluating long-term metabolic outcomes using survey-based data, including diabetes incidence and medication use. Together, Tasks 2 and 3 demonstrate the feasibility and value of integrating advanced exposure modeling techniques to capture both short-term and long-term pollutant variability relevant to different health endpoints.

Findings from the CHIS data showed significant positive associations between annual exposure to NO_2, $PM_{2.5}$ and O_3 and T2D incidence, indicating that exposure to these pollutants may contribute to diabetes onset. These associations were strongest in later diagnosis periods (2005–2015), corresponding with years of improved monitoring resolution and more consistent exposure patterns across California. In addition, long-term changes in atmospheric chemistry and climate-related factors, such as rising temperatures, increased photochemical activity, and more frequent stagnation events, may have contributed to higher or more biologically relevant O_3 exposures, thereby amplifying detectable health effects in more recent periods. The results on medication use revealed a consistent pattern of higher odds of diabetes medication use with increasing NO_2, $PM_{2.5}$ and O_3 levels, suggesting that exposure may worsen disease control or severity among diagnosed individuals. The magnitude of associations was greatest for $PM_{2.5}$, followed by O_3, with NO_2 showing smaller but steady effects. This gradient mirrors the differential toxicity of these pollutants and aligns with evidence from other large-scale studies linking fine particulates to systemic inflammation, insulin resistance, and disease progression.[11,12]

Results from the HCAI data further reinforce the acute impacts of air pollution on diabetes-related health care utilization. Both NO_2 and $PM_{2.5}$ exposures were significantly associated with elevated risks of T2D-related ED visits and hospital admissions. Among those hospitalized, exposure to these pollutants also correlated with longer LOS, reflecting higher clinical severity and potentially slower recovery. The acute effects appeared transient, with effect estimates

generally declining modestly from lag 0 to lag 3 days. This attenuation pattern suggests short-term, reversible physiological responses, such as systemic oxidative stress and inflammatory pathways, that may peak quickly after exposure. While some studies have reported stronger effects over 3 days of lagged effect,[12,154,155] our findings indicate that the largest effects occurred at lag 0, consistent with some research finding that pollutant exposure triggers near-immediate impacts on disease exacerbation or acute decompensation.[96,156,157] This pattern suggests that short-term air pollution exposure may induce transient, reversible physiological disturbances rather than sustained systemic damage. Acute elevations in NO_2 and $PM_{2.5}$ can trigger rapid-onset oxidative stress and pro-inflammatory responses, leading to endothelial dysfunction, altered glucose metabolism, and sympathetic activation within hours of exposure. These biological perturbations can exacerbate existing metabolic instability among individuals with diabetes, increasing the likelihood of emergency department visits or hospital admissions. The observed peak effect at lag 0 indicates that the body's response to pollution may occur almost immediately, consistent with experimental and epidemiologic studies showing that exposure to combustion-related pollutants quickly elevates circulating inflammatory markers (e.g., C-reactive protein, interleukin-6) and impairs insulin sensitivity.[158] Nonetheless, the gradual attenuation across days supports a sustained, though diminishing, influence over short time windows.

Further, across all stratified analyses, clear racial, ethnic, and linguistic disparities emerged in vulnerability to air pollution-related diabetes outcomes. Individuals classified in the "Other" race-ethnicity group consistently exhibited the highest susceptibility to NO_2, particularly for acute outcomes such as diabetes-related emergency department visits and hospitalizations, while Hispanic populations showed the strongest vulnerability to $PM_{2.5}$, especially in regions with higher particulate burdens such as Southern California and the Central Valley. Stratification by primary language further reinforced these inequities: individuals in the "Other-language" group experienced the largest and most rapidly increasing risks for both NO_2 and $PM_{2.5}$ over time, followed by Spanish speakers, whereas English speakers consistently showed the lowest effects. In contrast, non-Hispanic White populations demonstrated the lowest pollution-related risks across pollutants and endpoints. Together, these findings indicate that racially diverse, Hispanic, and linguistically isolated communities experience a disproportionate burden of air pollution–related diabetes risk, reflecting the combined influence of higher exposure levels, structural disadvantage, occupational risk, and reduced access to preventive and ongoing healthcare.

Analyses using CDPH mortality data revealed that both NO_2 and $PM_{2.5}$ were significantly associated with increased odds of T2D-related deaths, reinforcing the health burden of chronic air pollution exposure. $PM_{2.5}$ again showed the strongest effect estimates, consistent with its fine particulate nature and greater ability to penetrate deep into the lungs and systemic circulation. These findings are in line with numerous epidemiologic studies demonstrating robust

associations between particulate matter and all-cause or cause-specific mortality, including those related to metabolic and cardiovascular complications.[159–162]

Across all analyses involving acute effects, O_3 showed more complex and mixed effects. While some analyses showed modestly positive associations, others, displayed weaker or even negative effects.[23–26] Such inconsistencies are not uncommon in the literature and may be attributed to the spatial and temporal heterogeneity of O_3 formation, its inverse relationship with NO_2 in urban areas (due to titration by traffic emissions), and differential seasonal exposure patterns. Some prior studies have similarly reported null or inverse O_3 effects for metabolic outcomes,[163,164] whereas others observed positive associations, especially in regions with higher photochemical activity. The mixed findings in our study thus likely reflect both the complex chemistry of O_3 and the interplay of co-pollutant exposures.

In our study, the "Other" race-ethnicity category is not a single dominant group but rather a heterogeneous aggregation of multiple smaller populations whose individual sample sizes are insufficient for stable, independent estimation at the statewide level. This category primarily includes respondents identifying as American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, multiracial individuals, and smaller Asian subgroups not separately coded in the public or restricted-use data, as well as respondents with missing or unspecified race-ethnicity. Because each of these groups represents a relatively small fraction of the total population, further disaggregation would result in sparse data and unstable estimates, particularly for stratified and lagged analyses using case-crossover or survey-weighted models. The stronger effect estimates observed for the "Other" category therefore likely reflect aggregated vulnerability across multiple smaller and potentially underserved populations rather than the influence of a single dominant subgroup. We acknowledge this as an important limitation and note that future work with larger samples or targeted datasets would be needed to meaningfully disentangle heterogeneity within this category.

Taken together, these results highlight a coherent message across datasets: long-term exposure to NO_2 and $PM_{2.5}$ elevates the risk of developing T2D, increases the likelihood of disease progression requiring medical treatment and mortality, while acute exposures exacerbate complications leading to higher ED visits, hospital admissions, and prolonged LOS. The consistency of these associations across data sources, lag structures, and population subgroups underscores the robustness of the observed relationships.

This study further extends the air pollution diabetes literature by demonstrating robust associations between multiple ambient air toxics and both incident T2D and medication use. Effect estimates were internally consistent across lag structures and diagnosis periods. Nickel showed the strongest pooled association with incidence (OR≈1.64–1.66), with similarly elevated associations for medication use (OR≈1.22–1.32 across lags), indicating a potentially potent

metabolic effect. Benzene and 1,3-butadiene also demonstrated substantial increases in diabetes risk, particularly during 2000–2010 diagnosis periods, while lead and chromium exhibited stable, moderate associations.

This project also translated the exposure-response relationships identified in Task 5 into policy-relevant economic metrics, reinforcing the public health significance of the observed associations. By combining modeled concentration-response relationships with population exposure distributions, medical expenditure data, and established valuation approaches, this task provides a framework for estimating the economic benefits of reducing air pollution-related metabolic health burdens. Importantly, the economic analyses complement the epidemiologic findings by contextualizing health risks in terms directly relevant to regulatory decision-making, resource allocation, and cost-benefit considerations. These estimates underscore the substantial societal value of reducing individual pollutant exposures that are consistently associated with diabetes-related morbidity and mortality.

The robustness of our findings is supported by the exceptionally large, statewide datasets used in this study, encompassing millions of records across multiple independent health data sources. With sample sizes of this magnitude, alternative reasonable model specifications (e.g., different covariate sets, link functions, or lag structures) are expected to yield highly consistent results, as statistical uncertainty is substantially reduced and estimates are driven by stable population-level signals rather than sampling variability. In contrast, sensitivity of results to modeling choices reported in prior studies is largely attributable to smaller sample sizes, limited spatial coverage, or restricted temporal windows. Our analyses further demonstrate robustness through consistency across multiple health endpoints, pollutants, lag structures, and independent datasets, providing strong internal validation. This pattern aligns with the broader environmental epidemiology literature, which shows that when large population-based datasets and high-resolution exposure models are used, effect estimates are generally stable across alternative specifications, with differences primarily observed in studies with limited statistical power or localized samples.

From a policy and public health perspective, these findings emphasize the importance of continued air quality improvement efforts, especially targeting reductions in fine particulate and traffic-related pollutants. The disproportionate burden observed among minority and socioeconomically disadvantaged populations further highlights the need for equitable environmental health protections. Given the persistence of T2D as a major public health challenge, even modest pollutant-related increases in risk translate into substantial population-level impacts.

Simultaneous multi-pollutant health models were not implemented in this study because the primary objective was to quantify pollutant-specific concentration–response relationships rather

than cumulative effects. The large majority of epidemiologic studies evaluating air pollution and metabolic health outcomes rely on single-pollutant models, which remain the standard approach for estimating interpretable and policy-relevant effect sizes for individual pollutants. Multi-pollutant models, particularly those involving both criteria pollutants and air toxics, require specialized study designs to address collinearity, differential exposure error, and challenges in causal interpretation, and are more appropriately suited for cumulative impact exposure assessments. Because cumulative exposure assessment was not a stated objective of this project, and given the substantial additional methodological and computational demands required for valid multi-pollutant modeling, we focused on single-pollutant analyses in this work.

Despite the strengths of this study, including statewide coverage, large sample sizes, and high-resolution exposure modeling, several limitations warrant consideration. First, exposure assignment for diabetes-related emergency department visits and hospitalizations relied on ZIP code–level aggregation using population-weighted block group estimates. Although this approach captures small-area spatial variability, it may still misclassify individual-level exposures by not accounting for daily mobility, time–activity patterns, or indoor exposures. Such misclassification is expected to be largely nondifferential and would likely bias effect estimates toward the null.

Second, the observational design of this study precludes definitive causal inference. Although the case-crossover framework controls for time-invariant individual characteristics and lag-based analyses support temporally plausible exposure–response relationships, causal interpretation remains subject to the assumptions inherent in observational epidemiology.

Third, air toxics exposures were modeled at an annual temporal resolution and therefore could not be applied to short-term health outcomes such as emergency department visits, hospitalizations, or mortality analyzed using daily exposure windows. As a result, air toxics analyses were restricted to annual survey-based outcomes (diabetes incidence and medication use), limiting direct comparability with short-term criteria pollutant analyses.

Fourth, single-pollutant models were used for health analyses. While this approach is consistent with the majority of the air pollution epidemiology literature and facilitates interpretability, it does not explicitly quantify joint or cumulative effects of simultaneous multi-pollutant exposures. Multi-pollutant modeling requires specialized study designs and additional assumptions and was beyond the scope and resources of the current project. Future work will build on these results to address cumulative exposure and mixture effects.

Further, the ‘Other’ category in our analysis aggregated race-ethnicity groups not classified as Black, White, Asian, or Hispanic. This category may include groups such as Native American, Pacific Islander, Middle Eastern, multiracial individuals, and others. Potential differences within

these groups might influence the observed patterns and should be considered when interpreting the results.

Conclusion

This project provides a statewide evaluation of the impacts of ambient air pollution on T2D across the disease continuum, integrating high-resolution exposure modeling, large-scale epidemiologic analyses, and health economic valuation. Using multiple California-wide datasets, including CHIS incidence and medication use, HCAI hospital and emergency department records, and CDPH mortality data, this study offers a comprehensive assessment of how air pollution contributes to diabetes onset, disease management, acute complications, and mortality. The consistency of results across independent data sources strengthens confidence in the findings.

The exposure modeling framework captured fine-scale spatial and temporal variability in NO_2, $PM_{2.5}$, O_3, and selected air toxics, enabling population-level exposure assignment for millions of health records statewide. Despite limitations related to monitoring density for air toxics, the models demonstrated strong performance and provided a robust foundation for subsequent health analyses.

Across epidemiologic analyses, long-term exposure to NO_2, $PM_{2.5}$, and O_3 was associated with increased T2D incidence and greater use of diabetes medications, indicating a role for air pollution in both disease development and worsening disease control. $PM_{2.5}$ showed the strongest and most consistent associations across outcomes, while NO_2 exhibited smaller but stable effects. O_3 associations were more heterogeneous, likely reflecting its complex atmospheric chemistry and seasonal variability. This project also demonstrates that long-term exposure to multiple ambient air toxics is consistently associated with increased risk of T2D and greater diabetes medication use across California. These findings provide new evidence that hazardous air pollutants contribute meaningfully to California's diabetes burden.

Short-term exposure analyses using hospital and emergency department data showed that acute increases in NO_2 and $PM_{2.5}$ were associated with higher risks of T2D-related ED visits, hospital admissions, and longer hospital stays, suggesting that pollution can trigger clinically meaningful exacerbations. These effects were most pronounced shortly after exposure and attenuated over several days. Stratified analyses indicated that racial and ethnic minority populations often experienced greater pollution-related risks, highlighting persistent environmental health disparities.

Mortality analyses further demonstrated that higher annual exposures to $PM_{2.5}$ and NO_2 in the year preceding death were associated with increased diabetes-related mortality, supporting

the role of long-term air pollution exposure in disease progression and fatal outcomes. These findings are consistent with existing evidence linking particulate and traffic-related pollution to chronic cardiometabolic stress.

The health economic analyses translated these health impacts into substantial societal costs. Reductions in $PM_{2.5}$, NO_2, and O_3 were associated with large avoidable medical expenditures related to diabetes care, as well as significant reductions in the value of statistical lives lost due to diabetes-related mortality. $PM_{2.5}$ reductions yielded the greatest economic benefits, reflecting its strong influence across multiple health endpoints.

Overall, this integrated analysis demonstrates that ambient air pollution, particularly $PM_{2.5}$ and NO_2, contributes meaningfully to the onset, progression, and economic burden of T2D in California. These findings highlight air quality improvement as an effective and equitable strategy for reducing diabetes-related morbidity, mortality, and healthcare costs, and support the integration of air pollution control into chronic disease prevention and public health policy.

References

1. Vaidya V, Gangan N, Sheehan J. Impact of cardiovascular complications among patients with Type 2 diabetes mellitus: a systematic review. *Expert review of pharmacoeconomics & outcomes research*. 2015;15:487–497.

2. Usman MS, Khan MS, Butler J. The interplay between diabetes, cardiovascular disease, and kidney disease. 2021;

3. Henning RJ. Type-2 diabetes mellitus and cardiovascular disease. *Future cardiology*. 2018;14:491–509.

4. Sacchetta L, Chiriacò M, Nesti L, Leonetti S, Forotti G, Natali A, Solini A, Tricò D. Synergistic effect of chronic kidney disease, neuropathy, and retinopathy on all-cause mortality in type 1 and type 2 diabetes: a 21-year longitudinal study. *Cardiovascular Diabetology*. 2022;21:233.

5. Crawford AL, Laiteerapong N. Type 2 diabetes. *Annals of internal medicine*. 2024;177:ITC81–ITC96.

6. Rao X, Patel P, Puett R, Rajagopalan S. Air pollution as a risk factor for type 2 diabetes. *Toxicological Sciences*. 2015;143:231–241.

7. Balti EV, Echouffo-Tcheugui JB, Yako YY, Kengne AP. Air pollution and risk of type 2 diabetes mellitus: a systematic review and meta-analysis. *Diabetes research and clinical practice*. 2014;106:161–172.

8. Meo SA, Memon AN, Sheikh SA, Rouq FA, Usmani A, Hassan A, Arain SA. Effect of environmental air pollution on type 2 diabetes mellitus. *European Review for Medical & Pharmacological Sciences*. 2015;19.

9. Renzi M, Cerza F, Gariazzo C, Agabiti N, Cascini S, Di Domenicantonio R, Davoli M, Forastiere F, Cesaroni G. Air pollution and occurrence of type 2 diabetes in a large cohort study. *Environment international*. 2018;112:68–76.

10. Li Y, Xu L, Shan Z, Teng W, Han C. Association between air pollution and type 2 diabetes: an updated review of the literature. *Therapeutic advances in endocrinology and metabolism*. 2019;10:2042018819897046.

11. Rajagopalan S, Brook RD. Air pollution and type 2 diabetes: mechanistic insights. *Diabetes*. 2012;61:3037–3045.

12. O'Neill MS, Veves A, Sarnat JA, Zanobetti A, Gold DR, Economides PA, Horton ES, Schwartz J. Air pollution and inflammation in type 2 diabetes: a mechanism for susceptibility. *Occupational and environmental medicine*. 2007;64:373–379.

13. Lao XQ, Guo C, Chang L, Bo Y, Zhang Z, Chuang YC, Jiang WK, Lin C, Tam T, Lau AK. Long-term exposure to ambient fine particulate matter (PM2. 5) and incident type 2 diabetes: a longitudinal cohort study. *Diabetologia*. 2019;62:759–769.

14. Li S, Guo B, Jiang Y, Wang X, Chen L, Wang X, Chen T, Yang L, Silang Y, Hong F. Long-term exposure to ambient PM2. 5 and its components associated with diabetes: evidence from a large population-based cohort from China. *Diabetes Care*. 2023;46:111–119.

15. Wang Y, Shi L, Lee M, Liu P, Di Q, Zanobetti A, Schwartz JD. Long-term exposure to PM2. 5 and mortality among older adults in the southeastern US. *Epidemiology*. 2017;28:207–214.

16. Clark C, Sbihi H, Tamburic L, Brauer M, Frank LD, Davies HW. Association of long-term exposure to transportation noise and traffic-related air pollution with the incidence of diabetes: a prospective cohort study. *Environmental health perspectives*. 2017;125:087025.

17. Krämer U, Herder C, Sugiri D, Strassburger K, Schikowski T, Ranft U, Rathmann W. Traffic-related air pollution and incident type 2 diabetes: results from the SALIA cohort study. *Environmental health perspectives*. 2010;118:1273–1279.

18. Brook RD, Jerrett M, Brook JR, Bard RL, Finkelstein MM. The relationship between diabetes mellitus and traffic-related air pollution. *Journal of occupational and environmental medicine*. 2008;50:32–38.

19. Menzel DB. Ozone: an overview of its toxicity in man and animals. 1984;

20. Tan Q, Wang B, Ye Z, Mu G, Liu W, Nie X, Yu L, Zhou M, Chen W. Cross-sectional and longitudinal relationships between ozone exposure and glucose homeostasis: exploring the role of systemic inflammation and oxidative stress in a general Chinese urban population. *Environmental Pollution*. 2023;329:121711.

21. Zhong J, Allen K, Rao X, Ying Z, Braunstein Z, Kankanala SR, Xia C, Wang X, Bramble LA, Wagner JG. Repeated ozone exposure exacerbates insulin resistance and activates innate immune response in genetically susceptible mice. *Inhalation toxicology*. 2016;28:383–392.

22. Enweasor C, Flayer CH, Haczku A. Ozone-induced oxidative stress, neutrophilic airway inflammation, and glucocorticoid resistance in asthma. *Frontiers in Immunology*. 2021;12:631092.

23. Weaver AM, Bidulescu A, Wellenius GA, Hickson DA, Sims M, Vaidyanathan A, Wu W-C, Correa A, Wang Y. Associations between air pollution indicators and prevalent and incident diabetes in an African American cohort, the Jackson Heart Study. *Environmental Epidemiology*. 2021;5:e140.

24. McAlexander TP, Ryan V, Uddin J, Kanchi R, Thorpe L, Schwartz BS, Carson A, Rolka DB, Adhikari S, Pollak J. Associations between PM2. 5 and O3 exposures and new onset type 2 diabetes in regional and national samples in the United States. *Environmental research*. 2023;239:117248.

25. Hvidtfeldt UA, Sørensen M, Geels C, Ketzel M, Khan J, Tjønneland A, Overvad K, Brandt J, Raaschou-Nielsen O. Long-term residential exposure to PM2. 5, PM10, black carbon, NO2, and ozone and mortality in a Danish cohort. *Environment international*. 2019;123:265–272.

26. LaKind JS, Burns CJ, Pottenger LH, Naiman DQ, Goodman JE, Marchitti SA. Does ozone inhalation cause adverse metabolic effects in humans? A systematic review. *Critical Reviews in Toxicology*. 2021;51:467–508.

27. Hu M, Hao X, Zhang Y, Sun X, Zhang M, Zhao J, Wang Q. Long-term exposure to particulate air pollution associated with the progression of type 2 diabetes mellitus in China: effect size and urban–rural disparities. *BMC Public Health*. 2025;25:1565.

28. Weinmayr G, Hennig F, Fuks K, Nonnemacher M, Jakobs H, Möhlenkamp S, Erbel R, Jöckel K-H, Hoffmann B, Moebus S. Long-term exposure to fine particulate matter and incidence of type 2 diabetes mellitus in a cohort study: effects of total and traffic-specific air pollution. *Environmental Health*. 2015;14:53.

29. Qing S, Liang Z, Liang Y, Zhang R, Chen X, Wang W, Xu C, Lin F, Wang Y. Independent and combined relationships between nighttime light exposure, air pollution, PM2. 5 constituents, greenness and diabetes or high blood sugar: a national prospective cohort study. *BMC Public Health*. 2025;25:2755.

30. Su JG, Shahriary E, Sage E, Jacobsen J, Park K, Mohegh A. Development of over 30-years of high spatiotemporal resolution air pollution models and surfaces for California. *Environ Int*. 2024;193:109100.

31. Magliano DJ, Boyko EJ. IDF diabetes atlas. 2021;

32. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. *Diabetes research and clinical practice*. 2019;157:107843.

33. Gwira JA, Fryar CD, Gu Q. Prevalence of total, diagnosed, and undiagnosed diabetes in adults: united States, August 2021–August 2023. In: NCHS Data Briefs [Internet]. National Center for Health Statistics (US); 2024.

34. Inoue K, Liu M, Aggarwal R, Marinacci LX, Wadhera RK. Prevalence and control of diabetes among US adults, 2013 to 2023. *JAMA*. 2025;333:1255–1257.

35. Hu M, Le MH, Yeo YH, Wijarnpreecha K, Likhitsup A, Kim D, Chen VL. Diabetes prevalence and management patterns in US adults, 2001–2023. *Acta Diabetologica*. 2025;1–12.

36. CDC. National Diabetes Statistics Report [Internet]. Diabetes. 2024 [cited 2025 Oct 21];Available from: https://www.cdc.gov/diabetes/php/data-research/index.html

37. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, Stein C, Basit A, Chan JC, Mbanya JC. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes research and clinical practice*. 2022;183:109119.

38. Ong KL, Stafford LK, McLaughlin SA, Boyko EJ, Vollset SE, Smith AE, Dalton BE, Duprey J, Cruz JA, Hagins H, et al. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*. 2023;402:203–234.

39. Accili D, Deng Z, Liu Q. Insulin resistance in type 2 diabetes mellitus. *Nature Reviews Endocrinology*. 2025;1–14.

40. Khin PP, Lee JH, Jun H-S. Pancreatic beta-cell dysfunction in type 2 diabetes. *European Journal of Inflammation*. 2023;21:1721727X231154152.

41. Tomic D, Shaw JE, Magliano DJ. The burden and risks of emerging complications of diabetes mellitus. *Nature Reviews Endocrinology*. 2022;18:525–539.

42. Rajagopalan S, Brook RD, Salerno PR, Bourges-Sevenier B, Landrigan P, Nieuwenhuijsen MJ, Munzel T, Deo SV, Al-Kindi S. Air pollution exposure and cardiometabolic risk. *The Lancet Diabetes & Endocrinology*. 2024;

43. Rajagopalan S, Vergara-Martel A, Zhong J, Khraishah H, Kosiborod M, Neeland IJ, Dazard J-E, Chen Z, Munzel T, Brook RD. The urban environment and cardiometabolic health. *Circulation*. 2024;149:1298–1314.

44. Beulens JW, Pinho MG, Abreu TC, den Braver NR, Lam TM, Huss A, Vlaanderen J, Sonnenschein T, Siddiqui NZ, Yuan Z. Environmental risk factors of type 2 diabetes—an exposome approach. *Diabetologia*. 2022;65:263–274.

45. Misra BB, Misra A. The chemical exposome of type 2 diabetes mellitus: Opportunities and challenges in the omics era. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2020;14:23–38.

46. Association AD. The burden of diabetes in California. American Diabetes Association. https://diabetes. org/sites/default/files …; 2023.

47. Jhawar M, Mendez-Luck CA, Yu H, Meng Y-Y. Diabetes prevalence rates among adults across California legislative districts. 2005;

48. Health Promotion and Disease Prevention Program - Diabetes [Internet]. [cited 2025 Oct 21];Available from: https://healthpolicy.ucla.edu/our-work/diabetes

49. Al-Kindi SG, Brook RD, Biswal S, Rajagopalan S. Environmental determinants of cardiovascular disease: lessons learned from air pollution. *Nat Rev Cardiol*. 2020;

50. AB 2588 Air Toxics “Hot Spots” | California Air Resources Board [Internet]. [cited 2025 Oct 21];Available from: https://ww2.arb.ca.gov/our-work/programs/ab-2588-air-toxics-hot-spots

51. Schraufnagel DE. The health effects of ultrafine particles. *Experimental & molecular medicine*. 2020;52:311–317.

52. Prata JC, Da Costa JP, Lopes I, Duarte AC, Rocha-Santos T. Environmental exposure to microplastics: An overview on possible human health effects. *Science of the total environment*. 2020;702:134455.

53. Mauderly JL, Chow JC. Health effects of organic aerosols. *Inhalation toxicology*. 2008;20:257–288.

54. Zou B, Wilson JG, Zhan FB, Zeng Y. Air pollution exposure assessment methods utilized in epidemiological studies. *Journal of Environmental Monitoring*. 2009;11:475–490.

55. Wei X, Chang N-B, Bai K, Gao W. Satellite remote sensing of aerosol optical depth: Advances, challenges, and perspectives. *Critical Reviews in Environmental Science and Technology*. 2020;50:1640–1725.

56. Bowe B, Xie Y, Li T, Yan Y, Xian H, Al-Aly Z. The 2016 global and national burden of diabetes mellitus attributable to PM2· 5 air pollution. *The Lancet Planetary Health*. 2018;2:e301–e312.

57. Coogan PF, White LF, Yu J, Burnett RT, Marshall JD, Seto E, Brook RD, Palmer JR, Rosenberg L, Jerrett M. Long term exposure to NO2 and diabetes incidence in the Black Women’s Health Study. *Environmental research*. 2016;148:360–366.

58. Bevan GH, Al-Kindi SG, Brook RD, Münzel T, Rajagopalan S. Arteriosclerosis, Thrombosis, and Vascular Biology Ambient Air Pollution and Atherosclerosis Insights Into Dose, Time, and Mechanisms. *Arterioscler Thromb Vasc Biol*. 2021;41:628–637.

59. Ye M, Yang J, Li J, Wang Y, Chen W, Zhu L, Wang T, Liu J, Geng D, Yu Z. Progress in Mechanisms, Pathways and Cohort Studies About the Effects of PM2.5 Exposure on the Central Nervous System. *Reviews Env.Contamination (formerly:Residue Reviews)*. 2023;261:7.

60. Rajagopalan S, Al-Kindi SG, Brook RD. Air Pollution and Cardiovascular Disease: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2018;72:2054–2070.

61. Zhou L, Xiong Y, Sera F, Vicedo-Cabrera AM, Abrutzky R, Guo Y, Tong S, Coelho M de SZS, Saldiva PHN, Lavigne E. Associations of ambient exposure to benzene, toluene, ethylbenzene, and xylene with daily mortality: a multicountry time-series study in 757 global locations. *The Lancet Planetary Health*. 2025;

62. Hill BG, Rood B, Ribble A, Haberzettl P. Fine particulate matter (PM2.5) inhalation-induced alterations in the plasma lipidome as promoters of vascular inflammation and insulin resistance. *Am J Physiol Heart Circ Physiol*. 2021;320:H1836–H1850.

63. Wang Y, Xing C, Cai B, Qiu W, Zhai J, Zeng Y, Zhang A, Shi S, Zhang Y, Yang X, et al. Impact of antioxidants on PM2.5 oxidative potential, radical level, and cytotoxicity. *Science of The Total Environment*. 2024;912:169555.

64. Miller MR, Raftis JB, Langrish JP, McLean SG, Samutrtai P, Connell SP, Wilson S, Vesey AT, Fokkens PHB, Boere AJF, et al. Inhaled Nanoparticles Accumulate at Sites of Vascular Disease. *ACS Nano*. 2017;11:4542–4552.

65. Duan X, Zhang X, Chen J, Xiao M, Zhao W, Liu S, Sui G. Association of PM2.5 with Insulin Resistance Signaling Pathways on a Microfluidic Liver–Kidney Microphysiological System (LK-MPS) Device. *Anal. Chem.* 2021;93:9835–9844.

66. Ribble A, Hellmann J, Conklin DJ, Bhatnagar A, Haberzettl P. Fine particulate matter (PM2. 5)-induced pulmonary oxidative stress contributes to increases in glucose intolerance and insulin resistance in a mouse model of circadian dyssynchrony. *Science of The Total Environment*. 2023;877:162934.

67. Rajagopalan S, Park B, Palanivel R, Vinayachandran V, Deiuliis JA, Gangwar RS, Das L, Yin J, Choi Y, Al-Kindi S. Metabolic effects of air pollution exposure and reversibility. *The Journal of clinical investigation*. 2020;130:6034–6040.

68. Liviero F, Pavanello S. Epidemiological and mechanistic links between PM2. 5 exposure and type 2 diabetes: focus on the TRPV1 receptor. *Frontiers in Endocrinology*. 2025;16:1653375.

69. Liu F, Chen G, Huo W, Wang C, Liu S, Li N, Mao S, Hou Y, Lu Y, Xiang H. Associations between long-term exposure to ambient air pollution and risk of type 2 diabetes mellitus: a systematic review and meta-analysis. *Environmental pollution*. 2019;252:1235–1245.

70. Wang B, Xu D, Jing Z, Liu D, Yan S, Wang Y. Effect of long-term exposure to air pollution on type 2 diabetes mellitus risk: a systemic review and meta-analysis of cohort studies. *European journal of endocrinology*. 2014;171.

71. Yang M, Cheng H, Shen C, Liu J, Zhang H, Cao J, Ding R. Effects of long-term exposure to air pollution on the incidence of type 2 diabetes mellitus: a meta-analysis of cohort studies. *Environmental Science and Pollution Research*. 2020;27:798–811.

72. Yang B-Y, Fan S, Thiering E, Seissler J, Nowak D, Dong G-H, Heinrich J. Ambient air pollution and diabetes: a systematic review and meta-analysis. *Environmental research*. 2020;180:108817.

73. Wang B, Xu D, Jing Z, Liu D, Yan S, Wang Y. Mechanisms in endocrinology: effect of long-term exposure to air pollution on type 2 diabetes mellitus risk: a systemic review and

meta-analysis of cohort studies. *European Journal of Endocrinology*. 2014;171:R173–R182.

74. Eze IC, Hemkens LG, Bucher HC, Hoffmann B, Schindler C, Künzli N, Schikowski T, Probst-Hensch NM. Association between ambient air pollution and diabetes mellitus in Europe and North America: systematic review and meta-analysis. *Environmental health perspectives*. 2015;123:381–389.

75. Janghorbani M, Momeni F, Mansourian M. Systematic review and metaanalysis of air pollution exposure and risk of diabetes. *European journal of epidemiology*. 2014;29:231–242.

76. Azizi S, Dehghani MH, Nabizadeh R. Ambient air fine particulate matter (PM10 and PM2.5) and risk of type 2 diabetes mellitus and mechanisms of effects: a global systematic review and meta-analysis. *International Journal of Environmental Health Research* [Internet]. 2025 [cited 2025 Sept 30];Available from: https://www.tandfonline.com/doi/abs/10.1080/09603123.2024.2391993

77. Yu S, Zhang M, Zhu J, Yang X, Bigambo FM, Snijders AM, Wang X, Hu W, Lv W, Xia Y. The effect of ambient ozone exposure on three types of diabetes: a meta-analysis. *Environ Health*. 2023;22:32.

78. Balti EV, Echouffo-Tcheugui JB, Yako YY, Kengne AP. Air pollution and risk of type 2 diabetes mellitus: a systematic review and meta-analysis. *Diabetes Res Clin Pract*. 2014;106:161–172.

79. Kliengchuay W, Wen B, Aye TS, Aung HW, Suwanmanee S, Tawatsupa B, Laor P, Kongpran J, Wongsantichon J, Xu R, et al. Effect modification of fine particulate matter (PM2.5) related hospital admissions by temperature in Thailand: A nationwide time-series study. *Environ Res*. 2025;277:121467.

80. Zaheer H, Lv S, Li Z, Wu Z, Lu F, Guo M, Tao L, Gao B, Wang X, Li X, et al. Association between short-term exposure to ambient PM2.5 and its components with hospital admissions for patients with coronary heart disease and comorbid diabetes mellitus in Beijing, China. *Environ Res*. 2025;269:120729.

81. Liu M, Li Z, Lu F, Guo M, Tao L, Liu M, Liu Y, Deginet A, Hu Y, Li Y, et al. Acute effect of particulate matter pollution on hospital admissions for cause-specific respiratory diseases among patients with and without type 2 diabetes in Beijing, China, from 2014 to 2020. *Ecotoxicol Environ Saf*. 2021;226:112794.

82. Luo H, Liu C, Chen X, Lei J, Zhu Y, Zhou L, Gao Y, Meng X, Kan H, Xuan J, et al. Ambient air pollution and hospitalization for type 2 diabetes in China: A nationwide, individual-level case-crossover study. *Environ Res*. 2023;216:114596.

83. Gu J, Shi Y, Zhu Y, Chen N, Wang H, Zhang Z, Chen T. Ambient air pollution and cause-specific risk of hospital admission in China: A nationwide time-series study. *PLoS Med*. 2020;17:e1003188.

84. Song J, Liu Y, Zheng L, Gui L, Zhao X, Xu D, Wu W. Acute effects of air pollution on type II diabetes mellitus hospitalization in Shijiazhuang, China. *Environ Sci Pollut Res Int*. 2018;25:30151–30159.

85. Li J, Liu C, Cheng Y, Guo S, Sun Q, Kan L, Chen R, Kan H, Bai H, Cao J. Association between ambient particulate matter air pollution and ST-elevation myocardial infarction: A case-crossover study in a Chinese city. *Chemosphere*. 2019;219:724–729.

86. Li X, Tang K, Jin X-R, Xiang Y, Xu J, Yang L-L, Wang N, Li Y-F, Ji A-L, Zhou L-X, et al. Short-term air pollution exposure is associated with hospital length of stay and hospitalization costs among inpatients with type 2 diabetes: a hospital-based study. *J Toxicol Environ Health A*. 2018;81:819–829.

87. Ye Y, Ma H, Dong J, Wang J. Association between short-term ambient air pollutants and type 2 diabetes outpatient visits: a time series study in Lanzhou, China. *Environ Sci Process Impacts*. 2024;26:778–790.

88. Yao J, Brauer M, Wei J, McGrail KM, Johnston FH, Henderson SB. Sub-Daily Exposure to Fine Particulate Matter and Ambulance Dispatches during Wildfire Seasons: A Case-Crossover Study in British Columbia, Canada. *Environ Health Perspect*. 2020;128:67006.

89. Yin P, Luo H, Gao Y, Liu W, Shi S, Li X, Meng X, Kan H, Zhou M, Li G, et al. Criteria air pollutants and diabetes mortality classified by different subtypes and complications: A nationwide, case-crossover study. *J Hazard Mater*. 2023;460:132412.

90. Wu C, Yan Y, Chen X, Gong J, Guo Y, Zhao Y, Yang N, Dai J, Zhang F, Xiang H. Short-term exposure to ambient air pollution and type 2 diabetes mortality: A population-based time series study. *Environ Pollut*. 2021;289:117886.

91. Zhang Y, Xu R, Huang W, Morawska L, Johnston FH, Abramson M, Knibbs L, Matus P, Ye T, Yu W, et al. Short-term Exposure to Wildfire-Specific PM2.5 and Diabetes Hospitalization: A Study in Multiple Countries and Territories. *Diabetes Care*. 2024;47:1664–1672.

92. Zhang W, Zhang R, Tian T, Liu T, Dong J, Ruan Y. Acute effects of air pollution on type II diabetes mellitus hospitalization in Lanzhou, China. *Environ Geochem Health*. 2023;45:5927–5941.

93. Dzhambov AM, Dikova K, Georgieva T, Panev TI, Mukhtarov P, Dimitrova R. Short-term effects of air pollution on hospital admissions for cardiovascular diseases and diabetes mellitus in Sofia, Bulgaria (2009-2018). *Arh Hig Rada Toksikol*. 2023;74:48–60.

94. Gariazzo C, Renzi M, Marinaccio A, Michelozzi P, Massari S, Silibello C, Carlino G, Rossi PG, Maio S, Viegi G, et al. Association between short-term exposure to air pollutants and cause-specific daily mortality in Italy. A nationwide analysis. *Environ Res*. 2023;216:114676.

95. Liu X, Li Z, Guo M, Zhang J, Tao L, Xu X, Deginet A, Lu F, Luo Y, Liu M, et al. Acute effect of particulate matter pollution on hospital admissions for stroke among patients with type 2 diabetes in Beijing, China, from 2014 to 2018. *Ecotoxicol Environ Saf*. 2021;217:112201.

96. Kim H, Kim W, Choi JE, Kim C, Sohn J. Short-term Effect of Ambient Air Pollution on Emergency Department Visits for Diabetic Coma in Seoul, Korea. *J Prev Med Public Health*. 2018;51:265–274.

97. Yang J, Zhou M, Zhang F, Yin P, Wang B, Guo Y, Tong S, Wang H, Zhang C, Sun Q, et al. Diabetes mortality burden attributable to short-term effect of PM10 in China. *Environ Sci Pollut Res Int*. 2020;27:18784–18792.

98. Kan H, Jia J, Chen B. The association of daily diabetes mortality and outdoor air pollution in Shanghai, China. *J Environ Health*. 2004;67:21–26.

99. To T, Feldman L, Simatovic J, Gershon AS, Dell S, Su J, Foty R, Licskai C. Health risk of air pollution on people living with major chronic diseases: a Canadian population-based study. *BMJ Open*. 2015;5:e009075.

100. Yin P, Luo H, Gao Y, Liu W, Shi S, Li X, Meng X, Kan H, Zhou M, Li G, et al. Criteria air pollutants and diabetes mortality classified by different subtypes and complications: A nationwide, case-crossover study. *J Hazard Mater*. 2023;460:132412.

101. Alshaarawy O, Zhu M, Ducatman AM, Conway B, Andrew ME. Urinary polycyclic aromatic hydrocarbon biomarkers and diabetes mellitus. *Occupational and Environmental Medicine*. 2014;71:437–441.

102. Stallings-Smith S, Mease A, Johnson TM, Arikawa AY. Exploring the association between polycyclic aromatic hydrocarbons and diabetes among adults in the United States. *Environmental research*. 2018;166:588–594.

103. Khosravipour M, Khosravipour H. The association between urinary metabolites of polycyclic aromatic hydrocarbons and diabetes: A systematic review and meta-analysis study. *Chemosphere*. 2020;247:125680.

104. Mallah MA, Basnet TB, Ali M, Xie F, Li X, Feng F, Wang W, Shang P, Zhang Q. Association between urinary polycyclic aromatic hydrocarbon metabolites and diabetes mellitus among the US population: a cross-sectional study. *International Health*. 2023;15:161–170.

105. Yang L, Yan K, Zeng D, Lai X, Chen X, Fang Q, Guo H, Wu T, Zhang X. Association of polycyclic aromatic hydrocarbons metabolites and risk of diabetes in coke oven workers. *Environmental Pollution*. 2017;223:305–310.

106. Yang L, Zhou Y, Sun H, Lai H, Liu C, Yan K, Yuan J, Wu T, Chen W, Zhang X. Dose-response relationship between polycyclic aromatic hydrocarbon metabolites and risk of diabetes in the general Chinese population. *Environmental Pollution*. 2014;195:24–30.

107. Henriksen GL, Ketchum NS, Michalek JE, Swaby JA. Serum dioxin and diabetes mellitus in veterans of Operation Ranch Hand. *Epidemiology*. 1997;8:252–258.

108. Steenland K, Calvert G, Ketchum N, Michalek J. Dioxin and diabetes mellitus: an analysis of the combined NIOSH and Ranch Hand data. *Occupational and Environmental Medicine*. 2001;58:641–648.

109. Huang C-Y, Wu C-L, Yang Y-C, Chang J-W, Kuo Y-C, Cheng Y-Y, Wu J-S, Lee C-C, Guo H-R. Association between dioxin and diabetes mellitus in an endemic area of exposure in Taiwan: a population-based study. *Medicine*. 2015;94:e1730.

110. Gang N, Van Allen K, Villeneuve PJ, MacDonald H, Bruin JE. Sex-specific associations between type 2 diabetes incidence and exposure to dioxin and dioxin-like pollutants: A meta-analysis. *Frontiers in Toxicology*. 2022;3:685840.

111. Remillard RB, Bunce NJ. Linking dioxins to diabetes: epidemiology and biologic plausibility. *Environmental Health Perspectives*. 2002;110:853.

112. De Tata V. Association of dioxin and other persistent organic pollutants (POPs) with diabetes: epidemiological evidence and new mechanisms of beta cell dysfunction. *International journal of molecular sciences*. 2014;15:7787–7811.

113. Liang R, Feng X, Shi D, Yu L, Yang M, Zhou M, Zhang Y, Wang B, Chen W. Associations of urinary 1, 3-butadiene metabolite with glucose homeostasis, prediabetes, and diabetes in the US general population: role of alkaline phosphatase. *Environmental Research*. 2023;222:115355.

114. Zhou Z-L, Li J-Y, Wang W, Wang X-Q, Yin L, Li N-N, Li X-Z. Exposome-wide association study of environmental toxicant exposure and insulin resistance: Findings from US National Health and Nutrition Examination Survey. *Ecotoxicology and Environmental Safety*. 2026;309:119594.

115. Zhang H, Han Y, Qiu X, Wang Y, Li W, Liu J, Chen X, Li R, Xu F, Chen W. Association of internal exposure to polycyclic aromatic hydrocarbons with inflammation and oxidative stress in prediabetic and healthy individuals. *Chemosphere*. 2020;253:126748.

116. Su JG, Meng Y-Y, Chen X, Molitor J, Yue D, Jerrett M. Predicting differential improvements in annual pollutant concentrations and exposures for regulatory policy assessment. *Environment international*. 2020;143:105942.

117. Kanaroglou PS, Jerrett M, Morrison J, Beckerman B, Arain MA, Gilbert NL, Brook JR. Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmospheric Environment*. 2005;39:2399–2409.

118. Yang L, Jin S, Danielson P, Homer C, Gass L, Bender SM, Case A, Costello C, Dewitz J, Fry J. A new generation of the United States National Land Cover Database:

Requirements, research priorities, design, and implementation strategies. *ISPRS journal of photogrammetry and remote sensing*. 2018;146:108–123.

119. Lunetta RS, Knight JF, Ediriwickrema J, Lyon JG, Worthy LD. Land-cover change detection using multi-temporal MODIS NDVI data. In: Geospatial Information Handbook for Water Resources and Watershed Management, Volume II. CRC Press; 2022. p. 65–88.

120. Abatzoglou JT. Development of gridded surface meteorological data for ecological applications and modelling. *International journal of climatology*. 2013;33:121–131.

121. Levelt PF, Joiner J, Tamminen J, Veefkind JP, Bhartia PK, Stein Zweers DC, Duncan BN, Streets DG, Eskes H, van der A R. The Ozone Monitoring Instrument: overview of 14 years in space. *Atmospheric Chemistry and Physics*. 2018;18:5699–5745.

122. Zhang H, Lyapustin A, Wang Y, Kondragunta S, Laszlo I, Ciren P, Hoff RM. A multi-angle aerosol optical depth retrieval algorithm for geostationary satellite data over the United States. *Atmospheric Chemistry and Physics*. 2011;11:11977–11991.

123. Homer C, Dewitz J, Yang L, Jin S, Danielson P, Xian G, Coulston J, Herold N, Wickham J, Megown K. Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*. 2015;81:345–354.

124. Keeley JE, Syphard AD. Large California wildfires: 2020 fires in historical context. *Fire Ecology*. 2021;17:1–11.

125. Brown PT, Hanley H, Mahesh A, Reed C, Strenfel SJ, Davis SJ, Kochanski AK, Clements CB. Climate warming increases extreme daily wildfire growth risk in California. *Nature*. 2023;621:760–766.

126. Li S, Banerjee T. Spatial and temporal pattern of wildfires in California from 2000 to 2019. *Scientific reports*. 2021;11:8779.

127. Larsen LC, Sacramento CA. The ozone weekend effect in California: evidence supporting NOx emission reductions. *CARB report, http://www. arb. ca. gov*. 2003;

128. Felegari S, Sharifi A, Khosravi M, Sabanov S, Tariq A, Karuppannan S. Using Sentinel-2 data to estimate the concentration of heavy metals caused by industrial activities in Ust-Kamenogorsk, Northeastern Kazakhstan. *Heliyon*. 2023;9.

129. Barendregt JJ, Veerman JL. Categorical versus continuous risk factors and the calculation of potential impact fractions. *Journal of Epidemiology & Community Health*. 2010;64:209–212.

130. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *American journal of public health*. 1998;88:15–19.

131. Lin C-K, Chen S-T. Estimation and application of population attributable fraction in ecological studies. *Environmental Health*. 2019;18:52.

132. Viera AJ. Odds ratios and risk ratios: what's the difference and why does it matter? *Southern medical journal*. 2008;101:730–734.

133. Madeira F, Corda M, Martins C, Perelman J. Burden and economic impact of PM2. 5 exposure on Acute Myocardial Infarction in Portugal, 2011-2021. *European Journal of Public Health*. 2024;34:ckae144-430.

134. Zhang H, You S, Zhang M, Liu D, Wang X, Ren J, Yu C. The impact of atmospheric pollutants on human health and economic loss assessment. *Atmosphere*. 2021;12:1628.

135. Xie Y, Li Z, Zhong H, Feng XL, Lu P, Xu Z, Guo T, Si Y, Wang J, Chen L. Short-Term ambient particulate air pollution and hospitalization expenditures of cause-specific cardiorespiratory diseases in China: a multicity analysis. *The Lancet Regional Health–Western Pacific*. 2021;15.

136. Wang X, Yu C, Zhang Y, Shi F, Meng R, Yu Y. Attributable risk and economic cost of cardiovascular hospital admissions due to ambient particulate matter in Wuhan, China. *International Journal of Environmental Research and Public Health*. 2020;17:5453.

137. Jiang W, Chen H, Liao J, Yang X, Yang B, Zhang Y, Pan X, Lian L, Yang L. The short-term effects and burden of particle air pollution on hospitalization for coronary heart disease: a time-stratified case-crossover study in Sichuan, China. *Environmental Health*. 2022;21:19.

138. Zhou W, Wen Z, Peng W, Wang X, Yang M, Wang W, Wei J, Xiong H. Association of ambient particulate matter with hospital admissions, length of hospital stay, and hospital costs due to cardiovascular disease: time-series analysis based on data from the Shanghai Medical Insurance System from 2016 to 2019. *Environmental Sciences Europe*. 2023;35:46.

139. Ma Y, Zhang Y, Wang W, Qin P, Li H, Jiao H, Wei J. Estimation of health risk and economic loss attributable to PM2. 5 and O3 pollution in Jilin Province, China. *Scientific Reports*. 2023;13:17717.

140. Dieleman JL, Beauchamp M, Crosby SW, DeJarnatt D, Johnson EK, Lescinsky H, McHugh T, Pollock I, Sahu M, Swart V. Tracking US health care spending by health condition and county. *JAMA*. 2025;333:1051–1061.

141. Peterson C, Aslam MV, Niolon PH, Bacon S, Bellis MA, Mercy JA, Florence C. Economic Burden of Health Conditions Associated With Adverse Childhood Experiences Among US Adults. *JAMA Netw Open*. 2023;6:e2346323.

142. Institute for Health Metrics and Evaluation. United States Health Care Spending by Health Condition and County 2010-2019. 2025;

143. Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. *International journal of epidemiology*. 2013;42:860–869.

144. Shahar E, Shahar DJ. Causal diagrams and the logic of matched case-control studies. *Clinical Epidemiology*. 2012;137–144.

145. Wan F, Sutcliffe S, Zhang J, Small D. Does matching introduce confounding or selection bias into the matched case-control design? *Observational Studies*. 2024;10:1–9.

146. Greenland S, Pearl J, Robins JM. Confounding and collapsibility in causal inference. *Statistical science*. 1999;14:29–46.

147. Kenah E. A potential outcomes approach to selection bias. *Epidemiology*. 2023;34:865–872.

148. King G, Zeng L. Estimating risk and rate levels, ratios and differences in case-control studies. *Statistics in medicine*. 2002;21:1409–1427.

149. Cheung YB, Ma X, Lam K, Li J, Milligan P. Bias control in the analysis of case–control studies with incidence density sampling. *International Journal of Epidemiology*. 2019;48:1981–1991.

150. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *bmj*. 2016;352.

151. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *American journal of epidemiology*. 1985;122:904–914.

152. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *American journal of public health*. 1998;88:15–19.

153. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics*. 1993;865–872.

154. Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *American journal of epidemiology*. 1988;128:1185–1197.

155. Du N, Ji A-L, Liu X-L, Tan C-L, Huang X-L, Xiao H, Zhou Y-M, Tang E-J, Hu Y-G, Yao T, et al. Association between short-term ambient nitrogen dioxide and type 2 diabetes outpatient visits: A large hospital-based study. *Environ Res*. 2022;215:114395.

156. Zhang W, Zhang R, Tian T, Liu T, Dong J, Ruan Y. Acute effects of air pollution on type II diabetes mellitus hospitalization in Lanzhou, China. *Environmental geochemistry and health*. 2023;45:5927–5941.

157. Goldberg MS, Burnett RT, Yale J-F, Valois M-F, Brook JR. Associations between ambient air pollution and daily mortality among persons with diabetes and cardiovascular disease. *Environ Res*. 2006;100:255–267.

158. Pereira Filho MA, Pereira LAA, Arbex FF, Arbex M, Conceição GM, Santos UP, Lopes AC, Saldiva PHN, Braga ALF, Cendon S. Effect of air pollution on diabetes and cardiovascular diseases in São Paulo, Brazil. *Brazilian Journal of Medical and Biological Research*. 2008;41:526–532.

159. Pergoli LW. SUSCEPTIBILITY TO PARTICULATE MATTER AND HEALTH EFFECTS MEDIATED BY MICRORNAS CARRIED IN PLASMA EXTRACELLULAR VESICLES. 2015;

160. Li W, Tian A, Shi Y, Chen B, Ji R, Ge J, Su X, Pu B, Lei L, Ma R. Associations of long-term fine particulate matter exposure with all-cause and cause-specific mortality: results from the ChinaHEART project. *The Lancet Regional Health–Western Pacific*. 2023;41.

161. Wallwork RS, Colicino E, Zhong J, Kloog I, Coull BA, Vokonas P, Schwartz JD, Baccarelli AA. Ambient fine particulate matter, outdoor temperature, and risk of metabolic syndrome. *American journal of epidemiology*. 2017;185:30–39.

162. Jaganathan S, Jaacks LM, Magsumbol M, Walia GK, Sieber NL, Shivasankar R, Dhillon PK, Hameed SS, Schwartz J, Prabhakaran D. Association of long-term exposure to fine particulate matter and cardio-metabolic diseases in low-and middle-income countries: a systematic review. *International journal of environmental research and public health*. 2019;16:2541.

163. Mohammadi MJ, Fouladi Dehaghi B, Mansourimoghadam S, Sharhani A, Amini P, Ghanbari S. Cardiovascular disease, mortality and exposure to particulate matter (PM): a systematic review and meta-analysis. *Reviews on environmental health*. 2024;39:141–149.

164. LaKind JS, Burns CJ, Pottenger LH, Naiman DQ, Goodman JE, Marchitti SA. Does ozone inhalation cause adverse metabolic effects in humans? A systematic review. *Critical Reviews in Toxicology*. 2021;51:467–508.

165. Zanobetti A, Schwartz J. Is there adaptation in the ozone mortality relationship: a multi-city case-crossover analysis. *Environmental Health*. 2008;7:22.